UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUIMICAS
Departamento de Bioquímica y Biología Molecular

TESIS DOCTORAL

Modelling and gene therapy for pyruvate kinase deficiency

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Sergio López Manzaneda

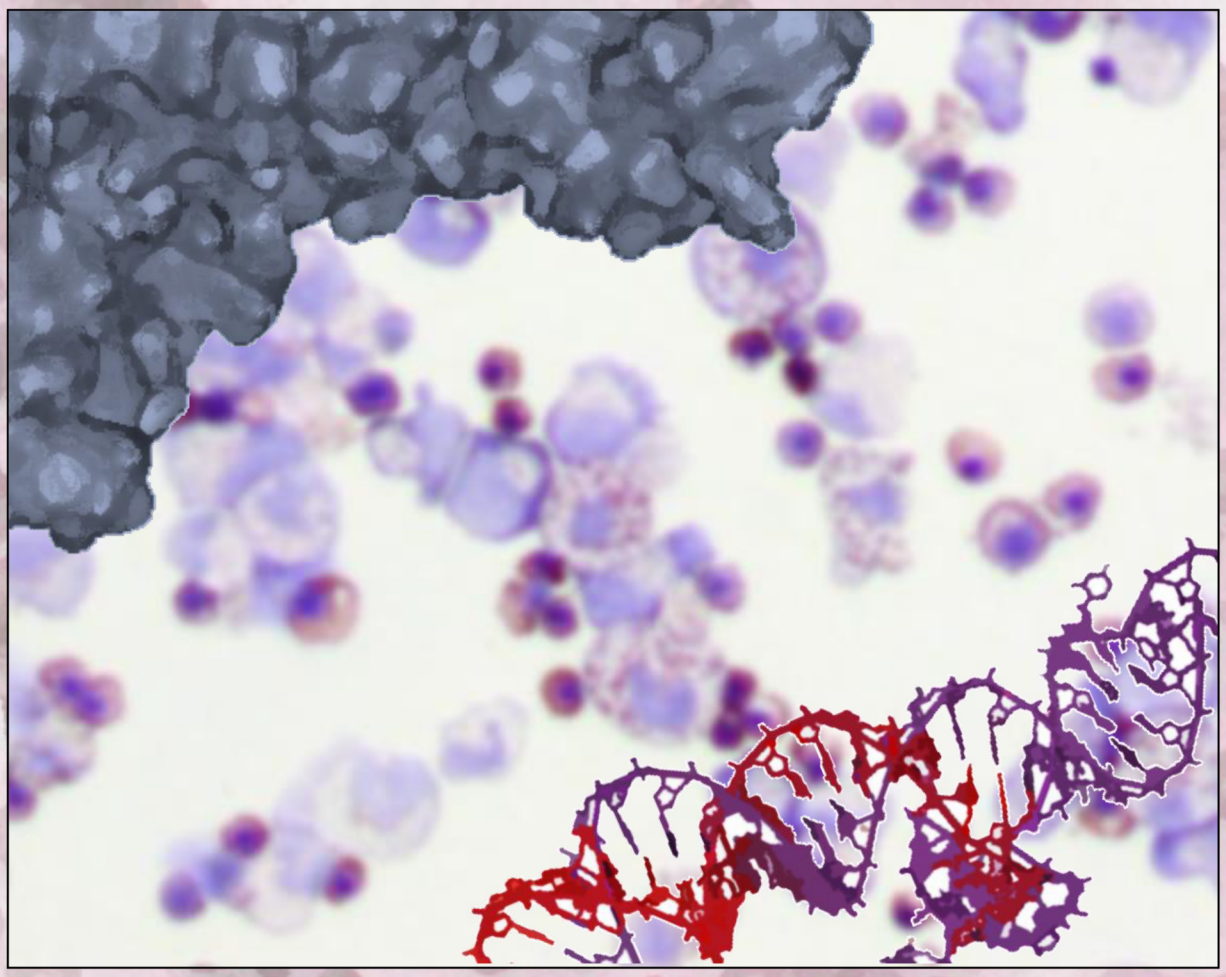
Director

José Carlos Segovia Sanz

Madrid
Ed. electrónica 2019

MODELLING AND GENE THERAPY FOR PYRUVATE KINASE DEFICIENCY

MODELIZACIÓN Y TERAPIA GÉNICA PARA LA DEFICIENCIA EN PIRUVATO QUINASA



Sergio López Manzaneda

UNIVERSIDAD COMPLUTENSE MADRID

Doctoral Thesis

-2018-

Director: José Carlos Segovia Sanz

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS QUÍMICAS
Departamento de Bioquímica y Biología Molecular

Universidad Complutense de Madrid

Facultad de Ciencias Químicas

Departamento de Bioquímica y Biología Molecular

DOCTORAL THESIS

MODELLING AND GENE THERAPY

FOR PYRUVATE KINASE DEFICIENCY

Sergio López Manzaneda

Madrid, 2018

Universidad Complutense de Madrid

Facultad de Ciencias Químicas

Departamento de Bioquímica y Biología Molecular

UNIVERSIDAD
COMPLUTENSE
MADRID

TESIS DOCTORAL

MODELIZACIÓN Y TERAPIA GÉNICA PARA LA DEFICIENCIA EN PIRUVATO QUINASA

Memoria presentada por Sergio López Manzaneda para optar al Grado de Doctor por la Universidad complutense de Madrid.

El trabajo experimental recogido en esta memoria ha sido realizado en la división de Terapias Innovadoras en el Sistema Hematopoyético del Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT) bajo la dirección del Doctor José Carlos Segovia Sanz.

Director de la tesis Doctoral:

Dr. José Carlos Segovia Sanz

Sergio López Manzaneda

Madrid, 2018

Agradecimientos

En primer lugar quiero agradecer esta tesis a todas las personas que, de una forma u otra, me han ayudado durante estos años a hacer de este trabajo lo que es hoy.

También quiero agradecerle a mi director, José Carlos Segovia, que apostase por mí. He disfrutado mucho estos años de nuestras conversaciones y brainstormings intentando averiguar el por qué de lo que salía y de lo que no. Gracias por darme la oportunidad de crecer como científico, por la paciencia y por los consejos.

Muchas gracias también a la gente del laboratorio en el que empecé, Chema, Antonio, Amalia, Isa, Patri, María, Marta y todos los demás. Y muchas gracias a ti Maga por convertirte en la mejor compi de aquella época. Si, chipi, ya sé que te debo una visita, dame tiempo que me lié con lo de la tesis.

Agradezco a toda la división de terapias innovadoras en el sistema hematopoyético su recibimiento y acogida. Desde el primer día me he sentido como en casa, quiero agradecer el apoyo y la ayuda que siempre me han ofrecido desde cada uno de los miembros de esta gran familia. Especialmente quiero agradecerle a la gente con la que he compartido la frustración científica cuando las cosas no salían: Paula, Susana, Óscar, María, Rosa y un largo etcétera. Por supuesto, mención especial ha de tener Guille, antes artista que científico. Gracias por enseñarles a todos los becarios lo que hay que tener para ser un buen *ciematero*.

Gracias también al gran equipo de gestión: Sole, Aurora, Mamen, Nadia y Sergio, la última resistencia ante la crisis. Gracias por vuestros malabarismos para que todos los reactivos/pagos/comisiones llegasen a tiempo. Como exministro de sondas y primers, sólo puedo daros las gracias por la fluidez de nuestra gestión conjunta. Mi más solemne y excelso agradecimiento.

Gracias a Rebeca y Omaira por ser, además de unas grandes personas, dos de las mejores profesionales del CIEMAT. Gracias por todo en lo que me habéis ayudado y por animarme siempre en cada sorting, a pesar de las cantidades ingentes de células que me tocaba manejar.

Gracias a los pequehematos clásicos: Diego, Bego, María, Fati, M.J., Miguel, Clara y todos los demás, y muy especialmente a mi María, oh! Gran Lideresa! Tus consejos me han permitido sobrevivir en la selva que es la tesis hasta el día de hoy, ¡mil gracias!

Gracias también a los pequehematos nuevos: Charito, Raqueliña, Carmensin (bueno, ella no es de hemato, es del mundo), Cristacho y su furia rociera, Sara, Fans y su amor por los corderitos, a Carapapa y su aún mayor amor por los moluscos, y todos los que se me olvidan. Gracias especialmente también a Aida por toda su ayuda y las conversaciones en las que arreglamos el país, y al gran equipo Andrea-Virgi. Virgi, eres una de las mejores

científicas que conozco y además salsera, ¿qué más se puede pedir? También le quiero agradecer a todos los pequeños hematos fugaces: Jordi, Julia, Inés, Vero murcianica, Cris Barcelona y demás, por entender el espíritu ciematero y enriquecerlo con vuestra personalidad.

Muchas gracias a los Epis presentes y pasados. Gracias a ti Epitelian Queen por enseñarnos lo que es el arte salsero más puro y por ser tan única, también a ti Vero, best singer ever, a Maesillo por tener esa personalidad culta e inspiradora y gracias también a ti adorable Hater, que te nos marchaste a la Europa del frio. Gracias también por presentarme a la gran Elena, antes conocida como la mae que baila, ahora como una buena amiga. Gracias a ti también morenita. También gracias al resto de epis fugaces, Flaco, Bailongo, Dianita y alguno que seguro me dejo sin mencionar.

Muchas gracias a todos los oncos, especialmente a la gran Carol, faro del frikismo, que aparte de hechizarnos para ir dos veces por semana a la capital de Europa, Móstoles, es capaz de dominar los ejércitos de mini-oncos para la hora del café. Gracias a todos los oncos por ser como sois: a Iris, a Sarita y su inquietud, Cris, Miri por sus anécdotas y su ayuda en las escape rooms, a la José por sus abrazos, Marta, Jesús y sus clases de baile en la cafetería, a la pitonisa Rake, Richi, Lau bailona e hija de moderdonia… en fin a todos. Ah, sí! Tres grandes pasaron por onco y se ganaron mi corazón, Bea tha Influencer, Vik Caracoli y la gran Chuchi, a vosotras también muchas gracias.

Y hablando de oncos…Muchas gracias a ti Canario, nadie ha creído más en mí que tú, ni si quiera yo, aún conservo un poco de criterio. No tengo palabras para agradecer todo lo que me has ayudado estos últimos años. Eres un gran científico y mejor persona. En esta tesis hay mucho de ti y lo sabes. Muchas gracias además por no dejarme desfallecer en esos momentos tan duros en los que te enfrentas a las grandes preguntas…"ya he hecho un 70 y 84…¿y ahora?… ¡Canario, rueda!"

Muchas gracias a ti Sokovi. Comenzamos este precioso viaje juntos, preguntándonos donde estaba el truco, y ahora, trucados perdidos, solo podemos pensar en la capacidad liberadora de un buen ochio de morcilla de caldera. No olvidaré todas las lecciones aprendidas contigo, el euskera mal, el marco incomparable, la conducción con Silvio, la anatomía con Romualdo y con Félix... bueno que te voy a contar de él. Dale duro a lo que te queda, nos vemos en Jaén siempre.

Muchas gracias a ti Yari, cuando me di cuenta que realmente no eras una doctora no di crédito, pero ahí estabas tú, running to the hill con todo el metal. Muchas gracias por las risas interminables, los viajes improvisados, haberme presentado a la familia Ignis y tu toda tu ayuda dentro y fuera del labo. Además, gracias por compartir conmigo la custodia de Jordi, esa oveja descarriada por el google maps que se quedó encerrada en ésta, la Cataluña del oeste.

Muchas gracias Bonafont, grande de España y enorme de la terreta. Pese a que con la edad las cosas no te ilusionen tanto, aun consigues inspirar con tu optimismo a todos cuanto te rodean. Estoy seguro que llegarás a ser un gran IP, DJ, Willy Fog o lo que te propongas porque tu ímpetu es imparable, te lo digo yo que he bailado contigo. Mil gracias guapete.

Muchas gracias súper Ester, tú me has enseñado que lo importante de caerse es saber caerse, reiteradamente y luego saber levantarse. Supe que eras una crack desde aquel torico en el que te conocí y me lo demuestras cada día, te sabes hacer querer. Y como dijo aquel, la vida es como una caja de bombones, nunca sabes cuándo ¡café y regalos!

Gracias por todo Patri, por venirte a lo loco al torico, por animarme desde que nos conocimos en aquella fresca isla e incluso por intentar reinsertar a Charlos. Sigue igual de loca siempre y no dejes que Charlos te venda nada…más. Si…gracias a ti también Charlos, ahora tengo un montón de productos inútiles en casa a muy buen precio…los que te conocemos sabemos que tienes unos pechos que no te caben en el corazón.

Gracias también a la única argentina muy española y muchos español. Gracias Maru por tu cariño y todas las historias relocas que me traés cada día. Sabés que mi sitio siempre está listo para uno más de tus kilombos.

Gracias también a las grandes gentes que conocí y con las que compartí grandes momentos durante la carrera: Ayuso, Antonio, Gong, Saru, Marina, Ana V, Jaime, e incluso, locamente, a Goma.

Eskerrik asko al eusko-team, Iker e Izaskun, que conocí en Glasgow, que me enseñaron un montón de palabras en ese precioso idioma y que hicieron de mi estancia una experiencia única.

Mila esker también a ti, olagarro. Nos conocimos de forma fortuita, y volvimos a coincidir cuando el mundo se dio la vuelta. Quizá fue el cambio de perspectiva, o la necesidad de equilibrio tras el fenómeno pero en cualquier caso, más que fortuito fue afortunado. Muchas gracias por mostrarme tu forma de sentir y de entender el mundo, me han cambiado y me han enseñado a entenderme.

Gracias sobre todo a mi familia más cercana, a mis padres, a mis tíos y muy especialmente a mis primos, que más que primos son hermanos, por hacer del cortijo el refugio que le da sentido al resto de la vida.

En ante-antepenúltimo lugar, y bajo amenaza, le agradezco a mi hermano Mario haberme engañado cuando era joven encendiendo mi curiosidad por la ciencia, explicándome las reacciones químicas en el bachillerato fingiendo que eran algo divertido. Menos mal que tu elegiste un oficio que realmente le da dinero a la familia que si no estaríamos perdidos…

Gracias a Javi, Almudena, Nico y sus respectivas multitud de primos/tíos por haberme tratado como uno más de la familia.

Por último he de agradecer a mi Ainho, mi vida, que has estado conmigo desde antes de ser quien soy, ayudándome a crecer, a pesar de que los primeros bocetos de mi persona estaban regular nada más.

PD: A ti, que creías que no te iba a mencionar porque me he olvidado de ti. Gracias por eso que tú y yo sabemos, y por aquello otro también.

Index

Index

Figure index

Table index

Summary

Pyruvate kinase deficiency (PKD) is the main cause of chronic non-spherocytic hemolytic anemia. This disease is a monogenic recessive disorder that occurs due to mutations in PKLR gene. Although this gene codifies the expression of two isoforms of pyruvate kinase, expressed in liver and in red blood cell s, only the erythroid lineage is affected. Pyruvate kinase is the enzyme that catalyzes the last step of the glycolysis, a critical exergonic reaction for the energetic survival of erythrocytes.

Due to its low prevalence (about 5:20,000), PKD is considered a rare disease. Severity of the disease is variable, ranging from mild cases, having hemoglobin levels around 11 g/dL and requiring blood transfusions just during exacerbations, to values below 8.5 g/dL in the most severe cases, being transfusion-dependent and fatal in the worst cases. Most severe cases are transfusion-dependent even after splenectomy, the most agresive palliative treatment.

Apart from palliative treatments, such as blood transfusions or splenectomy, the only existing curative treatment is the allogenic bone marrow transplantation. But this strategy has important limitations, such as the lack of histocompatible donors, the reject of the engraftment or the dangerous graft versus host disease. Gene therapy is a strategy that can solve all these limitations since it is based on the autologous transplantation of hematopoietic progenitors genetically corrected.

Regarding gene therapy approach, our group has shown PKD correction in a mouse model of the disease using a lentiviral-based gene therapy. The results of these pre-clinical trials allowed us to apply for orphan drug designation, obtaining it from EMA (European Medicines Agency) and from FDA (Food and Drug Administration) in 2014 and 2016, respectively.

The main objectives of this work were i) to continue studying the extent of hematopoietic alterations in PKD patient samples obtained from the physicians, especially regarding erythroid lineage; ii) more directly associated with the gene therapy program, to explore the limiting dose of corrected cells capable of rescuing the PKD phenotype in a mouse model; and iii) due to the difficulties to obtain PKD hematopoietic progenitors in

which to test gene therapy tools, the aim was the generation of PKD-like cells from healthy hematopoietic progenitors using two different approaches, mRNA disruption based on short hairpin RNAs and direct genome editing based on CRISPR/Cas9 system.

A complete analysis of peripheral blood samples from PKD patients, through the complete blood count and flow cytometry immunophenotyping, was performed. Erythroid lineage of patients was the most affected subset, showing an increase in erythroid progenitor population and in reticulocytes, a clear signal of the response of the bone marrow to the anemic stress. A deeper study to clarify the possible affectation of primitive hematopoietic progenitors in this disease was carried out based on colony forming unit assays and flow cytometry of $CD34^+$ marker. There was a small but not significant increase of hematopoietic progenitors in peripheral blood of PKD patients. Also, PK deficient $CD34^+$ hematopoietic progenitor cells from patients did not show impairments generating these colonies in semi solid media, meaning that these progenitors were not distinguishable from healthy progenitors. Furthermore, data collected from patients' relatives, which had loss of one PKLR functional allele, showed a decrease of erythroid parameters but not enough to reach an anemic situation. Nevertheless, the number of relatives studied was low, so these results must be carefully interpreted.

Regarding the limiting dose of cells corrected with our therapeutic lentiviral vector necessary to correct deficient features, conditioned PKD mice were transplanted with PKD hematopoietic progenitors transduced at different doses of viral vector. The number of transgene copies present in peripheral blood was evaluated and correlated with the hematologic profile. PKD phenotype was corrected when the number of transgene copies per cell was higher than 0.3 in mice peripheral blood. These results allow us to set a therapeutic threshold above which transduction efficacies must aim to ensure the feasibility of the gene therapy treatment of PKD patients in the clinical trial scenario.

The generation of human PKD-like cells that recapitulate the disease features was carried out by shRNAs and CRISPR/Cas9 system.

With respect to the shRNA approach, hematopoietic progenitors from healthy donors were transduced with shRNA lentiviral vectors and differentiated in vitro towards erythroid lineage. The use of non-transduced and shRNA transduced cells in the same well

allowed us to observe an alteration in normal cell growth of the shRNA transduced cells, which also occurred with the control shRNA. Nevertheless, sorting-based selection of transduced cells permitted the enrichment of this population and the confirmation of the model by the evaluation of the decrease of pyruvate kinase activity.

In the same way, hematopoietic progenitors from healthy donors were modified by CRISPR/Cas9 technology. Different locations of PKLR gene, first exons 8, 9 and 11, and afterwards different regions from exon 9 to 10, were selected to be altered by Cas9 nuclease action. These hematopoietic progenitors were electroporated with Cas9 nuclease and specific guideRNAs for the regions previously commented. The use of exon 9 as target for Cas9 action completely eliminated RPK protein production, showing the importance of this region for generation of the enzyme. To control the outcomes of the genome editing process, guides were used in couples, generating a predictable precise deletion pattern as the most represented event. These human PKD-like cells were differentiated in vitro towards erythroid lineage, demonstrating, finally, the decrease of pyruvate kinase activity in these mature PKD-like cells, not finding alterations along in vitro erythroid differentiation and being, therefore, an appropriated and novel model to test new advanced therapies for PKD.

Resumen

La deficiencia en piruvato quinasa (PKD, por sus siglas en inglés) es la principal causa de anemia hemolítica no esferocítica. Es una enfermedad monogénica-recesiva que se origina por mutaciones en el gen PKLR. Aunque este gen codifica la información para la producción de dos isoformas de piruvato quinasa, expresadas en hígado y en eritrocitos, solo el linaje eritroide se ve afectado. La piruvato quinasa es la enzima que cataliza la última reacción de la glicólisis, un proceso exergónico crucial para la supevivencia energética de los eritrocitos.

Dada su baja prevalencia (alrededor de 5 de cada 20.000 habitantes), PKD es considerada una enfermedad rara. La severidad de la enfermedad es variable, desde casos leves con niveles de hemoglobina de 11 g/dL que solo necesitan transfusiones de eritrocitos puntuales durante exacerbaciones de la enfermedad, hasta casos graves con niveles de hemoglobina inferiores a 8.5 g/dL, dependientes de transfusiones periódicas que incluso pueden causar la muerte. Los casos más severos son dependientes de transfusiones periódicas incluso después de ser esplenectomizados, el tratamiento paliativo utilizado más agresivo.

Al margen de los tratamientos paliativos, como transfusiones y la esplenmectomía, solo existe un tratamiento curativo, el trasplante alogénico de médula ósea. Pero este procedimiento tiene importantes limitaciones, como son la dificultad de encontrar donantes compatibles, la posibilidad de rechazo del injerto o la enfermedad de injerto contra huésped. La terapia génica es una estrategia que solventaría estas limitaciones ya que se basa en el trasplante autólogo de progenitores hematopoyéticos genéticamente corregidos.

Respecto a la aproximación de terapia génica, nuestro grupo demostró corrección de la enfermedad en un modelo múrido de PKD corrigiendo mediante el uso de un vector lentiviral. Estos ensayos preclínicos nos permitieron la aprobación del vector como medicamento huérfano por la las agencias europea y americana del medicamento (EMA; European Medicines Agency y FDA; Food and Drug Administration) en 2014 y 2016 respectivamente.

Los principales objetivos de este trabajo han sido i) continuar estudiando el alcance de las alteraciones hematopoyéticas en muestras de pacientes, especialmente en el linaje

eritroide; ii) más directamente relacionado con el programa de terapia génica, el objetivo fue averiguar la cantidad mínima de células corregidas capaz de rescatar el fenotipo PKD en un modelo de ratón; y iii) debido a las dificultades en la obtención de muestras de progenitores hematopoyéticos deficientes, el objetivo fue la generación de células deficientes en PK a partir de progenitores hematopoyéticos procedentes de donantes sanos mediante la utilización de dos aproximaciones diferentes, eliminación de ARN mensajero mediante shRNA, y edición directa del genoma mediante el sistema CRISPR/Cas9.

Se analizaron muestras de sangre periférica de los pacientes de PKD mediante hemograma e inmunofenotipo utilizando citometría de flujo. El linaje más afectado fue el eritroide, mostrando un gran incremento del número de progenitores eritroides y reticulocitos circulantes, una clara señal de la respuesta de la médula ósea al estrés producido por la situación anémica. Igualmente, se llevó a cabo un estudio más exhaustivo de la posible afectación de los progenitores hematopoyéticos más primitivos a través del ensayo de unidades formadoras de colonias y del marcaje de $CD34^+$ en las células circulantes. Estos experimentos mostraron un pequeño, aunque no significativo, aumento del número de células $CD34^+$ en el torrente sanguíneo de los pacientes con respecto a los donantes sanos. Además, los progenitores hematopoyéticos deficientes no mostraron peor capacidad de generación de colonias en cultivos semi-sólidos que sus homólogos de donantes sanos. Los valores obtenidos de la sangre periférica de familiares de pacientes, que solo tenían afectos uno de los alelos, estaban ligeramente disminuidos pero no lo suficiente como para llegar a una situación de anemia. En cualquier lugar, el número de muestras de familiares analizadas fue demasiado bajo por lo que estos resultados se han de tomar con cautela.

Respecto al número mínimo de células corregidas necesario para restaurar valores sanos en el modelo animal el planteamiento fue el siguiente. Ratones PKD acondicionado fueron trasplantados con progenitores hematopoyéticos que previamente habían sido transducidos con diferentes dosis del vector terapéutico. El número de copias de transgén fue evaluado en las células presentes en la sangre periférica y este resultado se correlacionó con el nivel de corrección de la enfermedad alcanzado en cada caso. El fenotipo deficiente fue corregido en los ratones que tenían al menos 0,3 copias de vector por célula. Este resultado nos ha permitido establecer un umbral por encima del cual deben permanecer los

niveles de transducción para hacer posible el tratamiento de PKD mediante terapia génica de cara a un futuro ensayo clínico.

La generación de células humanas PKD a partir de progenitores hematopoyéticos sanos se llevó a cabo mediante el uso de shRNA y el sistema CRISPR/Cas9.

En la aproximación con shRNA, progenitores hematopoyéticos de donantes sanos fueron transducidos con vectores lentivirales conteniendo shRNAs y fueron diferenciados hacia linaje eritroide. El cultivo conjunto de células transducidas y no transducidas nos permitió comprobar que los vectores utilizados ralentizaban el crecimiento de las células transducidas, incluidas las que lo habían sido con el vector control. En cualquier caso, la separación de las células transducidas mediante "cell sorting" permitió el enriquecimiento de la población rica en shRNA, así como la confirmación de estas células como deficientes en actividad piruvato quinasa mediante el análisis de actividad enzimática específico.

Un procedimiento similar se llevó a cabo en el caso del sistema CRISPR/Cas9. Diferentes regiones del gen *PKLR*, en un principio los exones 8, 9 y 11 y posteriormente regiones comprendidas entre los exones 9 y 10, fueron las elegidas para que actuara la nucleasa Cas9. Progenitores hematopoyéticos procedentes de donantes sanos fueron electroporados con la nucleasa y con las guías de RNA específicas para las regiones previamente comentadas. El uso del exón 9 como objetivo eliminó completamente la producción de la enzima, demostrando la importancia de dicha región para la proteína. Las guías fueron utilizadas en parejas para tener un mayor control sobre los eventos de reparación del ADN, ya que de este modo producían en su mayoría un patrón de eliminación de nucleótidos predecible. Estos progenitores modificados fueron diferenciados hacia linaje eritroide, demostrando finalmente un descenso de la actividad PK tras la edición del genoma y su capacidad para diferenciarse hasta los estadios más maduros. Estas propiedades hacen de estas células un modelo humano novedoso y útil en el que probar nuevas terapias para el tratamiento de PKD.

Abbreviations

293T	Human Embryonic Kidneys (HEK) Transformed With Large T Antigen.
2-PG	2-Phosphoglycerate
3-PG	3-Phosphoglycerate
aa	Amino Acid
ADP	Adenosine Diphosphate
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
BFU-E	Burst-Forming Unit-Erythroid
BPGM	Biphosphoglycerate Mutase
BSA	Bovine Serum Albumin
CFU-E	Colony-Forming Unit–Erythroid
CFU-GM	Colony Forming Unit Granulocyte-Macrophages
CMP	Common Myeloid Progenitor
CNSHA	Chronic Non-Spherocytic Hemolytic Anemia
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/ Cas9
CytB5R	NADH-Cytochrome B5 Reductase
DMEM	Dulbecco's Modified Eagle´S Medium
DSB	Double Strand Break
EF1α	Elongation Factor-1 Alpha
EMP	Erythoblast-Macrophage Protein
EryD	Definitive Erythroid Cells
EryP	Primitive Erythroid Cells
FBP	Fructose-1,6-Bisphosphate
FLI-1	Friend Leukemia Integration 1 Transcription Factor
G6P	Glucose-6-Phosphate
GATA-1	GATA binding protein 1
GMP	Granulocyte-Macrophage Progenitors
GSH	Glutathione

GSSG	Glutathione Disulphide
GVHD	Graft Versus Host Disease
H_2O_2	Hydrogen Peroxide
Hb	Hemoglobin
HK	Hexokinase
HSC	Hematopoietic Stem Cell
HSCT	Hematopoietic Allogeneic Stem Cell Transplantation
ICAM4	Intracellular Adhesion Molecule 4
IGF-1	Insulin-like Growth Factor-1
IMDM	Iscove´S Modified Dulbeco´S Medium
IMS	Immunomagnetic Separation
iPSCs	Induced Pluripotent Cells
KLF1	Kruppel Like Factor 1
LCL	Lymphoblastoid Cell Line
LPK	Liver Pyruvate Kinase
M	Missense Mutation
MCH	Mean Corpuscular Hemoglobin
MCV	Mean Corpuscular Volume
MEP	Megakaryocytic-Erythroid Progenitor (MEP)
mHb	Methemoglobin
MIT	Massachusetts Institute of Technology
MLP	Multipotent Lymphoid Progenitor
MM	Mismatches
MOI	Multiplicity of Infection
MPP	Multipotential Progenitor Cell
NADH	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NGS	Next Generation Sequencing
NM	Non-Missense Mutation
PAM	Protospacer Adjacent Motif
PBS	Phosphate-Buffered Saline
PEP	Phosphoenolpyruvate

PFK-1	Phosphofructokinase 1
PGI	Phosphoglucose Isomerase
PGK	Phosphoglycerate Kinase Promoter
PK	Pyruvate Kinase
PKD	Pyruvate Kinase Deficiency
PMCA	Plasma Membrane Ca^{2+}-ATPase
PU.1	PU-Box Binding Site, A Purine-Rich DNA Sequence
qPCR	Quantitative PCR
RBC	Red Blood Cells
ROC	Receiver Operating Characteristic
RPK	Red Blood Cell Pyruvate Kinase
RPMI	Roswell Park Memorial Institute
Scrb	Scramble
SFFV	Spleen Focus-Forming Virus
shRNA	Short Hairpin RNA
TALEN	Transcription Activator-Like Effector Nucleases
TFRC	Transferrin Receptor 1
TSS	Transcription Start Site
TU	Transducing Units
VCAM-1	Vascular Cell Adhesion Molecule-1
VCN	Vector Copy Number
Wpre	Woodchuck Hepatitis Virus Post-Transcriptional Regulatory Element

Introduction

Introduction

Along this thesis manuscript different subjects will be discussed, all of them having a rare disease in common, pyruvate kinase deficiency (PKD). PKD is produced by mutations in *PKLR* gene, which encodes liver and red blood cell isoforms of pyruvate kinase (LPK and RPK, respectively). The disease appears when mutations affect RPK activity which dramatically impairs red blood cell lifespan.

To be aware when RPK takes importance for erythroblast´s life, a general overview of erythropoiesis is about to be explained at the beginning of this introduction. Then, RPK will be introduced by a brief explanation about the protein structure, gene structure and gene regulation. To further visualize the problem that a defective RPK can be, a short review of the erythrocyte´s metabolic processes will be presented highlighting the steps in which pyruvate kinase (PK) acts as keystone of red blood cells (RBC) survival. Also, a review of the disease, diagnosis and available treatments, including future ones as gene therapy and gene editing, are going to be explained.

Hematopoiesis is a sequential process based on the self-renewal/differentiation equilibrium. All blood lineages arise from the same cell type, the hematopoietic stem cell. As long as cells acquire a more committed profile to a specific lineage, they lose their self-renewal capability and their chance of differentiate to other lineages (Figure 1). Hematopoiesis can be classified in two big lineages, lymphoid (that include B and T lymphocytes, natural killer cells) and myeloid (that includes granulocytes, platelets and erythrocytes). Other hematopoietic cells such as dendritic cells and monocytes can mature from common myeloid progenitor but also from myeloid-lymphoid progenitors, exemplifying the dynamic equilibrium among lineages.[1]

Erythropoiesis

1. Primitive and definitive erythropoiesis

There are two well differentiated types of erythropoiesis: primitive (EryP) and

definitive (EryD). Although both types share some basic features, they arise from different mesoderm populations of the embryo.[2]

In the first wave of hematopoiesis, detectable after the 3rd week of gestation, EryP provides oxygen to the embryo during its rapid growth phases. Primitive erythroid progenitors are different from the definitive counterpart. They are bigger and express different globin genes than the definitive ones.[3] EryP is produced in the yolk sac and egress to the circulation nucleated mostly composed by erythroid cells. Unlike in the case of reptilian and avian erythrocytes, mammalian EryP nuclei are eventually expelled as they circulate.[4]

In humans, the second hematopoietic wave begins in the yolk sac at week 4. This second wave is composed by erythroid and myeloid cells. Fetal liver hematopoiesis becomes active contributing to the already active yolk sac hematopoiesis by weeks 5-6. EryP and EryD coexist at this stage. Then, EryD population rises becoming the main source of hematopoietic progenitors, throughout the rest of the fetal life.[2]

Finally, in the third wave, the hematopoiesis migrates to seed the thymus and bone marrow reaching the final niche of mature hematopoiesis. At birth, EryD is produced in bone marrow being, hereby, the source of all hematopoiesis throughout the adult life-span.[2,5]

2. Erythroid differentiation in definitive erythropoiesis

Depending on the ability to self-renewal and to differentiate into one or multiple hematopoietic lineages, progenitors committed towards erythroid lineage have been classified as follows (Figure 1).

Multipotent progenitor (**MPP**) cells are primitive progenitors still capable of generating all type of hematopoietic cells. Although molecular changes that determine whether MPP cells become erythroid progenitor are not completely clear, it is known that an increase in the transcription factor PU.1 promotes the differentiation towards common myeloid progenitors (**CMP**) that have lost the ability to give rise to lymphoid cells. Then, zinc finger transcription factor GATA-1 makes possible the differentiation toward the erythroid lineage, reaching the megakaryocytic-erythroid progenitor (**MEP**) stage. Finally,

two transcription factors, KLF1 and FLI-1, compete to differentiate MEP into either erythroid or megakaryocyte cells, respectively.[6]

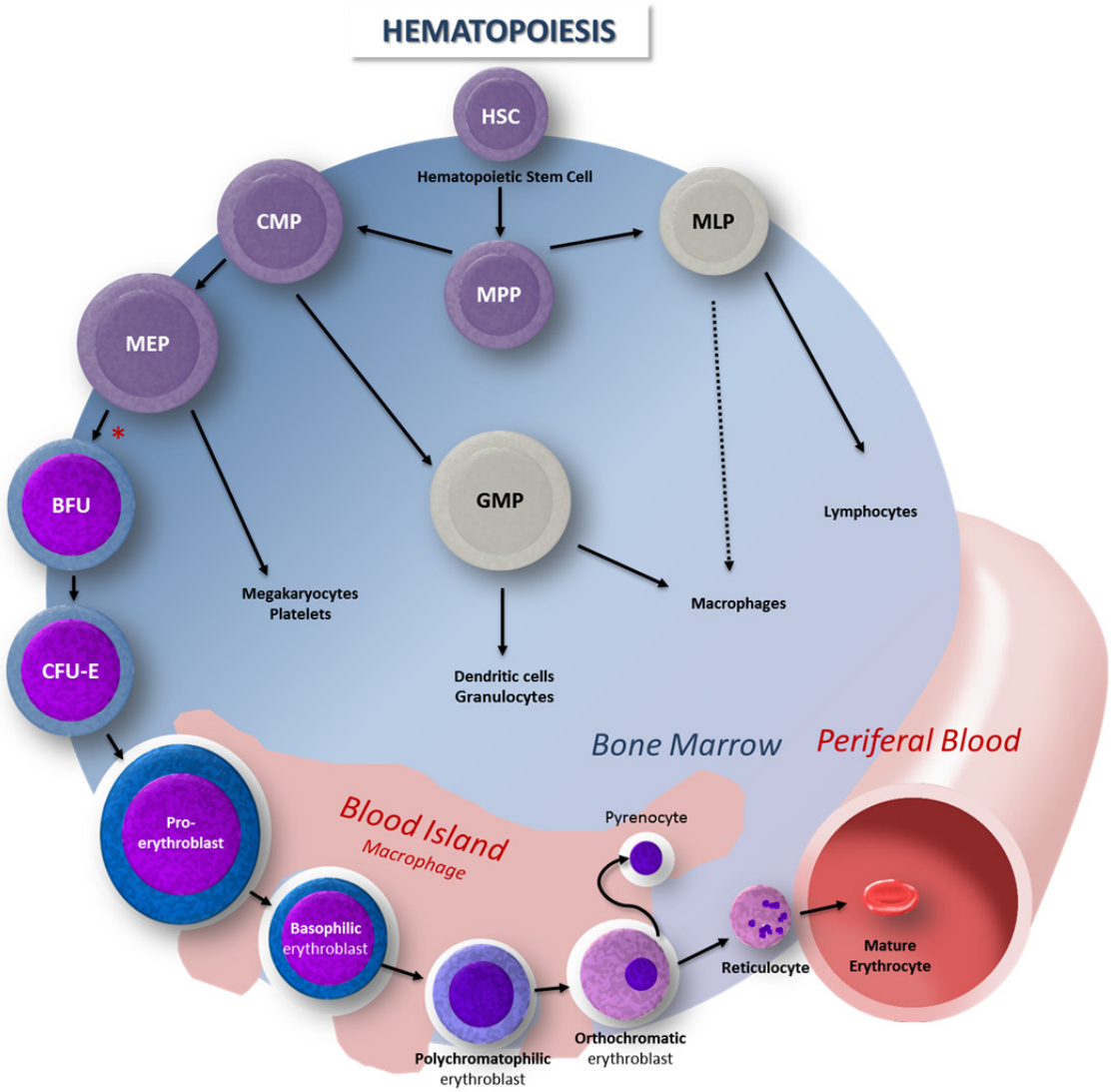


Figure 1. Hematopoiesis hierarchy.
Hematopoietic stem cell (HSC), multipotent progenitor (MPP), multipotent lymphoid progenitor (MLP), common myeloid progenitors (CMP), granulocyte-macrophage progenitors (GMP), megakaryocytic-erythroid progenitor (MEP). Red asterisk label the beginning of erythroid differentiation, which occurs mainly on blood islands.[1]

After MEP, eight different developmental stages along definitive erythroid maturation have been described. Similar to the above mentioned, all of them allocate also in the bone marrow except the two final cell types (reticulocytes and mature erythrocytes), that circulate in the blood stream.

The most immature progenitor of the erythroid specific compartment is called Burst-forming unit-erythroid (**BFU-E**). These progenitors receive their name from their behavior forming colonies in semisolid media. After 14 days of *in vitro* culture, human BFU-Es appear as colonies with a central core, surrounded by smaller satellite clusters in a "burst" shape. Transferrin receptor 1 (TFRC, also called CD71), necessary for transporting the iron which afterwards will be part of heme group, begins to be expressed at the end of this stage.[7,8]

Thereafter, the next identified differentiation step is the named Erythroid-Colony Forming Unit (**CFU-E**). They are also defined because of its *in vitro* growth in semi-solid media. Human CFU-Es are small erythroid colonies that are observable after 5-7 days in culture. Survival of this cell type is strictly dependent of erythropoietin (EPO) and high levels of this molecule induce differentiation to more committed steps.[3] CFU-Es are also capable to adhere to the macrophage surface. This property makes possible the formation of the structure where the next steps of erythroid differentiation take place, the so called blood island.[5,6]

The basic organization of a blood island is a central macrophage surrounded by as many as 30-48 erythroid committed cells.[6,9] This cell-cell contact is necessary for the correct maturation of the erythroid progenitors toward the erythroblast state. The first interaction occurs thanks to the action of erythroblast-macrophage protein (EMP). EMP is present on both erythroblasts and macrophages and is in charge of the initial joining. But not only EMP ensures macrophage-erythroblast interaction. Erythroid precursors start to express several adhesion proteins such as $\alpha_4\beta_1$ integrin and intracellular adhesion molecule 4 (ICAM4) which interact with vascular cell adhesion molecule-1 (VCAM-1) and α_V integrins respectively, present on macrophages.[10] The presence of $\alpha_4\beta_1$ and $\alpha_5\beta_1$ integrins decrease over maturation to make easy the egress of mature erythroblasts.[11] Other adhesion proteins found in central macrophages, such as CD169 and CD163, have been described to play a role in the blood islands.[10] For example, it has been studied that CD163 scavenges hemoglobin–haptoglobin complexes from the circulation, being important in hemoglobin homeostasis.[12]

These macrophages also have a key role in the last stage of maturation, being the cells that must help to remove and degrade the nuclei from orthochromatic erythroblasts. Macrophages recognize protein S bound to the phosphatidylserines present in the surface of

the nucleus, here called pyrenocyte, through their Mer tyrosine kinase receptor.[9] After this recognition, nuclei are phagocyted by the central macrophage turning orthochromatic erythroblast into reticulocytes. Macrophages, equipped with high levels of DNase II, recycle the ingested DNA and facilitate the enucleated egression of the reticulocyte to the peripheral blood.[4] It is known that this is a fast process that occurs within 10 minutes. Another important central macrophage function is the direct transfer of iron to the erythroblasts.[12]

Along their maturation in the blood islands, erythroblasts are classified as follows:

Pro-erythroblast. It has 20-25 μm of diameter and its nucleus occupies approximately 80% of cell area. Pro-erythroblasts present a slightly oval shape and have a characteristic intense basophilic cytoplasm due to the high polyribosome concentration. Moreover, ferritin and hemoglobin are already present at this stage.

Basophilic erythroblast. Along differentiation the cell size decreases. The mean diameter is 17 μm and their nuclei occupy about 75% of cell surface. It exhibits a dark violet chromatin with pink-stained clumps and like its predecessor shows a strong basophilic cytoplasm.

Polychromatophilic erythroblast. After mitotic division, basophilic erythroblast turns into polychromatophilic erythroblast. The new cell type suffers some changes: the diameter decreases until 13 μm, and its nucleus occupies less than 50% of the surface. The increase of hemoglobin decreases the polyribosome basophilia turning the cytoplasm in gray color. In this stage, CD71 expression reaches its maximum levels all over the maturation.[7] Additionally, polychromatophilic erythroblasts carry out the last mitotic division of the erythroid maturation.

Orthochromatic erythroblast. Cell diameter decreases to 10-15 μm and its nucleus occupies only a quarter of cell surface. Furthermore, nucleus shows a polarized distribution along the cytoplasm, preparing the enucleation. Orthochromatic erythroblast shows big motility capacity, probably necessary for nucleus ejection. Hemoglobin staining is predominant, and this is the last erythroblast bound to blood islands.

Reticulocyte. Orthochromatic erythroblast divides itself in reticulocyte and pyrenocyte (nucleus) helped by the macrophages. It receives its name because the rests of mitochondria, polyribosomes, Golgi apparatus and other organelles give it a reticular aspect. Reticulocyte matures in approximately 48-72 hours. During reticulocyte maturation, polyribosomes stop protein production and separate into monoribosomes. Cell size decreases, membrane structure is remodeled and gradually loses CD71. These changes make it more elastic and stable.

Mature erythrocyte. Finally, the mature erythrocyte with its classical biconcave shape arises out from the reticulocytes. It is the smallest erythroid cell, around 7.5 μm, and its membrane is prepared to be extremely deformable. This feature allows them to enter into capillaries of 2.8 μm diameter to transport oxygen anywhere. Their lifespan is around 100-120 days and its cell surface decreases over time. This decrease (up to 20% of its area) is related with cell ageing, producing a decrease in their deformability and, hence, an impairment of oxygen delivery. To maintain the normal amount of erythrocytes (5 x 10^6 RBC/μl), 2 x 10^{11} new erythrocytes need to be produce every day.[5,6,13]

Red blood cell pyruvate kinase

1. Protein structure and regulation

Pyruvate kinase catalyzes the final irreversible reaction of anaerobic glycolysis, also known as Embden–Meyerhof pathway, transforming phosphoenolpyruvate (PEP) into pyruvate. The energy released in this reaction is used to transform adenosine diphosphate (ADP) into adenosine triphosphate (ATP).[14] Pyruvate is the first non-phosphorylated intermediate and plays a central role in metabolism. PK activity requires two divalent cations, usually Mg^{2+}, and a monovalent cation, normally K^+.[15]

PK is an enzyme largely conserved through evolution. It has been characterized from a number of prokaryotes and eukaryotes. Nearly in all organisms, PK acts as homo-tetramer, having 500-600 amino acids (aa) each monomer, under allosteric regulation.[16] According with its allosteric regulation, prokaryote PK is classified in type I, activated by fructose-1, 6-bisphosphate (FBP) and inhibited by ATP, and type II, activated by adenosine

monophosphate (AMP) and mono-phosphorylated sugars. Eukaryote PK is allosterically activated by FBP and shows cooperative activity after PEP binding. Eukaryotic PK is more similar to the prokaryote PK type I.[17]

In human, PK is a step ahead in complexity. PK functions are carried out by four different isozymes, encoded by two different genes. Each of these isozymes is expressed under tissue specific promoters. *PKM* (EC: 2.7.1.40; 15q23) gene contains the sequences for two different isoforms, called M1 and M2. Alternative splicing of the *PKM* mRNA produces PKM1 expressed in adult tissues, such as muscle, and PKM2 expressed in rapidly proliferating tissues, such as fetal tissues.[14] PKM isoforms are tightly regulated and PKM2 is important in the oncology field because an increased PKM2 expression in different malignancies, like breast or colon cancers, has been described.[19] The other two isoforms are encoded by the *PKLR* gene. Although officially is called Pyruvate kinase, *PKLR* (EC: 2.7.1.40; 1q22) is also named Pyruvate kinase 1 (to distinguish it from *PKM*) or Pyruvate kinase L/R (referring to liver/Red Cell isoforms). Similarly to what happen in *PKM*, alternative splicing of *PKLR* mRNAs produces LPK isoform, expressed in liver, and RPK, expressed in red blood cells.

Each PK monomer presents 3 domains (Figure 2): A domain, $(\alpha/\beta)_8$ barrel; B domain, β-strand inserted between β_3 and α_3 of domain A; and C domain, (α+β) open-sheet. Eukaryotic PK contains also an N-terminal helical domain.[16]

Apart from tissue regulation, mammalian PKs show different regulation and kinetics. M1 is independent of allosteric modifiers and works following hyperbolic Michaelis-Menten kinetics. Meanwhile M2, LPK and RPK show sigmoidal kinetics, are allosterically activated by PEP and FBP, and inhibited by ATP.[16,18] LPK and RPK are generated as a homotetramer (62 kDa each monomer) and afterwards two out of the four monomers are partially proteolyzed in its N terminus region generating smaller monomers (57-58 kDa). Thus, the functional tetramer is formed by combination of two complete and two proteolyzed monomers. This cleavage is related with cell aging and seems to be important for enzyme activity.[20] Moreover, it has been described that LPK is also regulated by phosphorylation of a serine residue at the N terminus of the protein.[17]

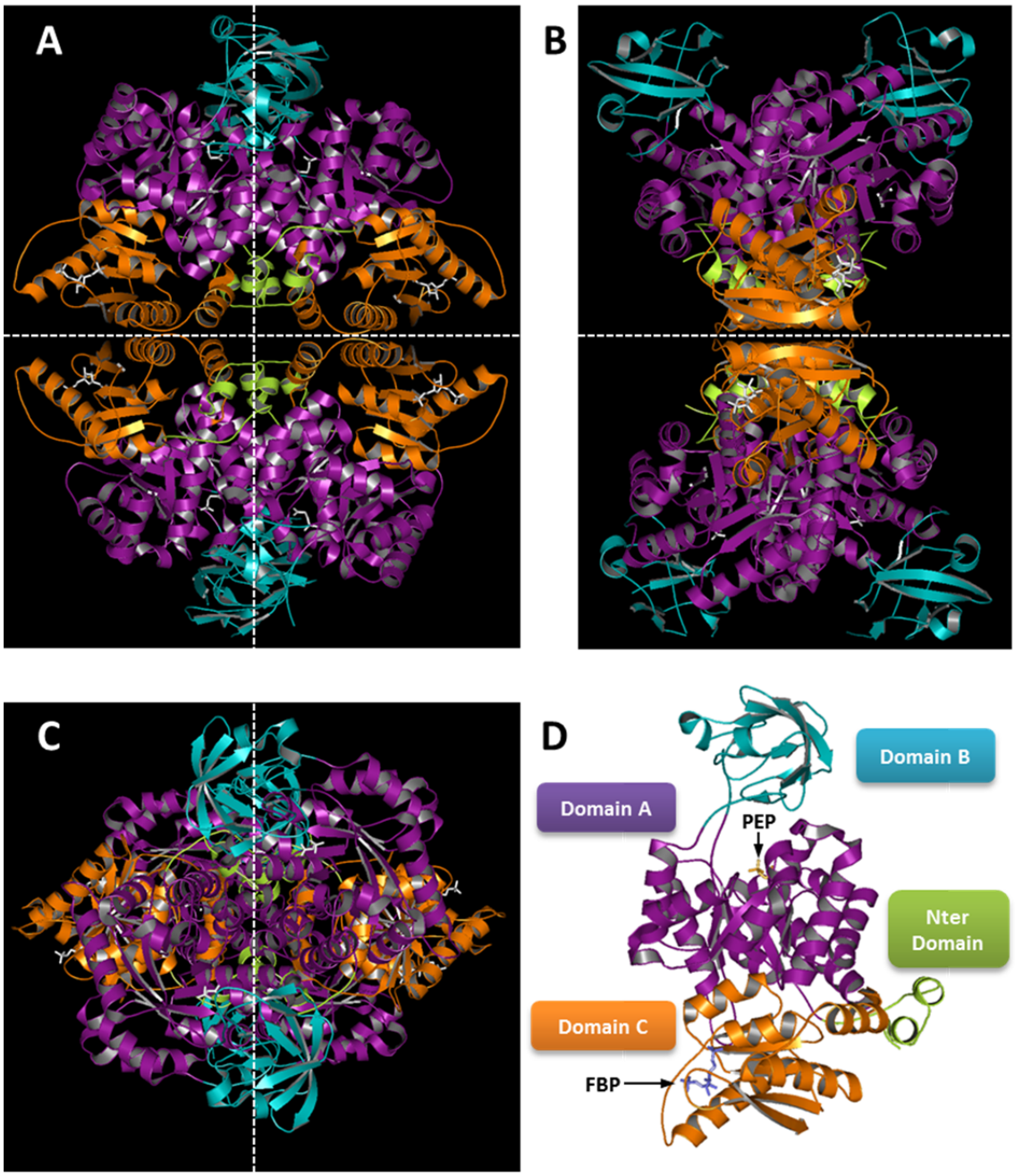


Figure 2. Pyruvate kinase structure.
(**A**, **B** and **C**) PK tetramer from different points of view: top-down perspective, 90 degrees vertically rotated and 90 degrees horizontally rotated, respectively. (**D**) PK monomer. PEP: bound to the catalytic site and FBP bound to the allosteric activator site. Color code represents protein domains. Nter: N-terminal domain. Model based on PDB: 2VGB [18]

Regarding the presence of the protein on erythroid cells, must be mention that PK activity is carried out by both PKM2 and RPK isoforms in different proportion along erythroid maturation. RPK starts to take importance at basophilic stage and progressively

substitutes PKM2 becoming the main isozyme from orthochromatic erythroblast to mature erythrocytes (Figure 3).[21]

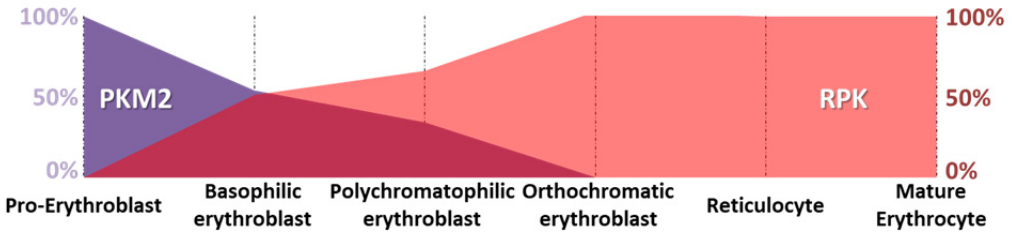


Figure 3. PKM2 vs RPK contribution to PK activity of PKM2/RPK isoforms along erythroid differentiation.

2. *PKLR*, the gene

As it has been explained, *PKLR* contains the genetic information for both LPK and RPK enzymes. Specific expression relies on the promoter regions (*PKLR* promoter region showed on Figure 4, A). It was back in 1987 when Noguchi first categorically showed that the two isozyme mRNAs are encoded by the same gene only differing in the 5'-terminal sequence.[22] *PKLR* is a 19.42 kb gene composed by 12 exons, which first two are specific for each isozyme (Figure 4, B). There are two accepted nomenclatures for the gene: naming the exons from 1 to 12, so exon 1 is erythroid and 2 hepatic; or naming from 1 to 11 for both isoforms, being 1(R) or 1(L) the first exon of RPK or LPK, respectively. Hereafter nomenclature used will be the 1-12 exons one. Erythroid exon number 1, is 14 times bigger than the liver homologue, number 2, which implies different protein lengths (574 aa for the RBC isoform and 543 aa for the liver isoform; Figure 4, C and D).[23]

The expression of both isozymes depends on the regions flanking their first exon. On rat (81% homology between human *PKLR* and rat *Pklr*), LPK expression is controlled by three regions (PKL-I, PKL-II and PKL-III) and a TATA box at intron 1. Therefore the transcribed region occurs from exon 2 to exon 12. [24] On the other hand, first RPK transcript includes liver exon which is finally excluded by alternative splicing process. RPK expression is regulated by the region 5´upstream its first exon. Classically, a large region, from -870 to -54 bp, has been considered important for its expression in the erythroid lineage.[25] Some authors consider the proximal (-120 bp) region as promoter site and the distal (until -150bp) as enhancer sequences.[26] Above all, some promoter and enhancer features have been found at 5' of exon 1. More precisely, four GATA motifs (at -73, -202, -226 and -248 bp), two

CAC/Sp1 boxes (-107 and-119 bp) and an enhancer region (PKR-RE, -87 bp) seem to play important roles in erythroid expression (Figure 4, A).[25,27,28]

The relevance of the proximal GATA motif has been corroborated in accordance with some PKD cases reported in which regions from -71 to -73 bp were mutated.[29–31] Another PKD case reported by Manco and Ribeiro showed mutations at -156 bp.[32] Some other mutations at nucleotides -83 and -109 (PKR-RE and CAC box, respectively) highlighted their crucial function in RPK expression.[27,33] Finally, *in vitro* mutations at -248 and -324 bp turned in a non-functional RPK.[27]

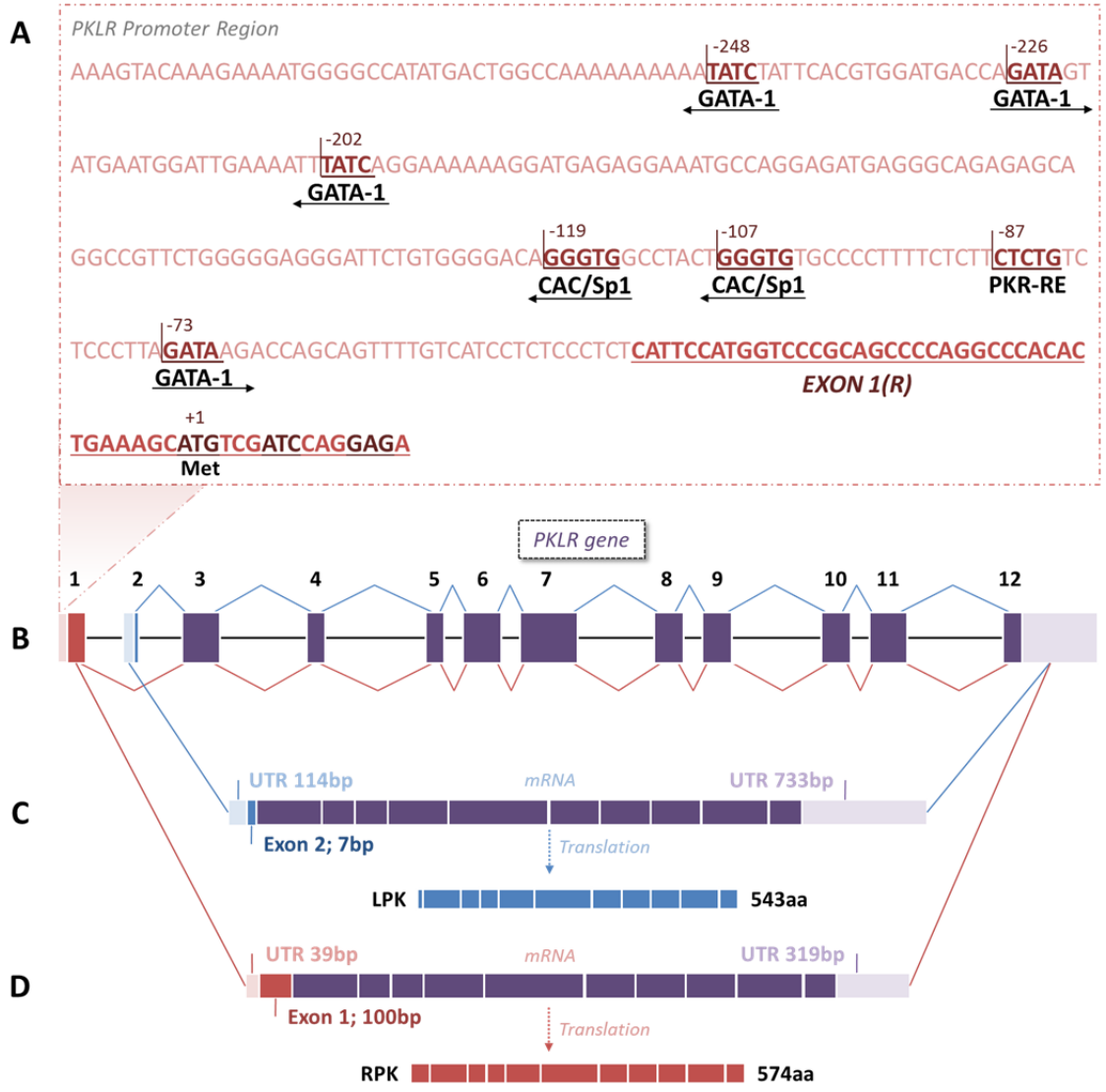


Figure 4. *PKLR* gene scheme, promoter, transcripts, isoforms.
(**A**) *PKLR* promoter region including binding sequences and distances with respect to transcription start site (TSS). (**B**) *PKLR* gene. Exons and UTR regions are represented by boxes: red, the erythroid exon; blue, the liver exon; purple, exons common for both isoforms; light red, erythroid isoform 5' UTR; light blue, liver isoform 5' UTR; light purple, 3' UTR common in both isoforms. Introns are represented as a black horizontal line. (**C**) mRNA and protein length of liver and (**D**) red blood cell isoforms.

3. Metabolic importance of PK in red blood cells

PKLR gene is expressed in both liver and red blood cells. Therefore, mutations in this gene affect PK activity of hepatocytes and erythrocytes. However, the differential nature of these cell types means different impact of these mutations. RPK activity is the main source of ATP in erythrocytes while hepatocytes have nucleus and mitochondria, and consequently, other ways to obtain ATP such as the use of other metabolic shunts based on mitochondrial functions or the overexpression of LPK, something that erythrocytes cannot do. This is why alterations on *PKLR* gene produce RBC deficiency without liver affection.

This lack of ATP production caused by mutations in *PKLR* gene dramatically alters all erythrocyte processes, affecting RBC lifespan in many ways, such as:

Glycolysis alterations. (Figure 5) PK deficiency blocks glycolysis. **PEP** cannot be transformed in pyruvate so it is accumulated in the RBC. This metabolite regulates negatively the activity of Glucose-6-phosphate isomerase (**GPI**; which converts glucose-6-phosphate in fructose-6-phosphate) and phosphofructokinase (**PFK-1;** a transferase which converts fructose-6-phosphate into fructose 1, 6-bisphosphate), that are the catalyzers of the second and the third step of glycolysis, respectively. Apart from this blocking at the beginning of the pathway, metabolites upstream PEP are also accumulated. Therefore, total amount of 2-phosphoglycerate (**2-PG**) and 3-phosphoglycerate (**3-PG**) increase. 3-PG activates allosterically biphosphoglycerate mutase (**BPGM**) which bypasses the direct conversion of 2-PG into 3-PG through the Rappaport-Leubering shunt. This pathway avoids the generation of ATP producing 2,3-biphosphoglycerate (**2,3-BPG**) instead, which is also accumulated in PK deficient RBC.[34–37]

Because of the accumulation of 2,3-BPG, some authors support the idea of the increase of Donnan Ratio as main cause of the hemolytic anemia. In other words, changes in glycolysis dynamics disturb the equilibrium of charged substances across the two sides of the membrane. Apparently, the increase (2-3 times the normal value) of 2, 3-BPG, due to its polyanionic character, and the decrease of intracellular KCl reduce the normal Donnan ratio. These changes would explain the alterations in the PK deficient erythrocyte volume producing echinocytosis.[38,39]

ATP depletion. Historically, ATP depletion has been considered the main cause of hemolytic anemia in red cell enzymopathies.[40] Normally, 90% of glucose is used to produce ATP trough anaerobic glycolysis.[41]

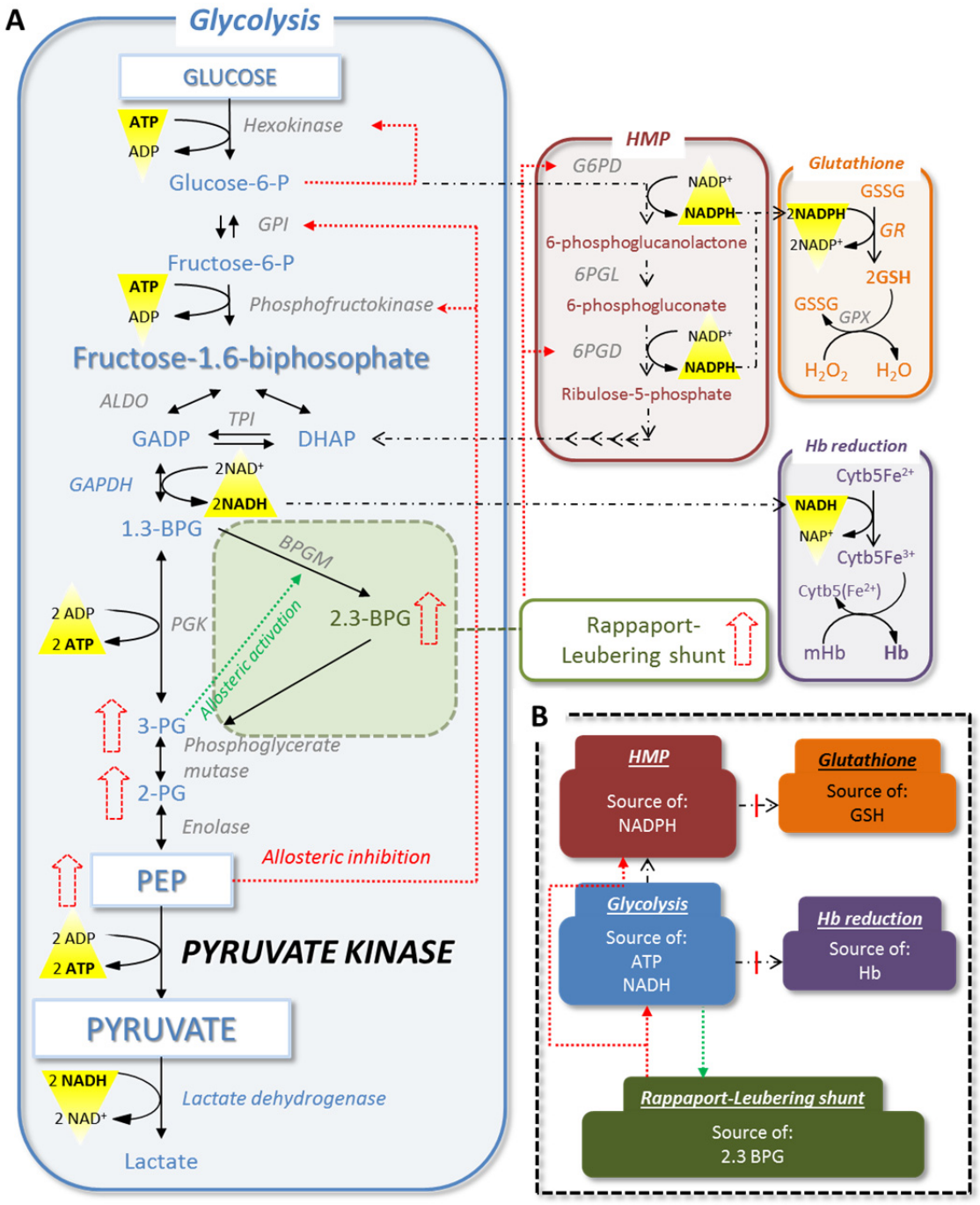


Figure 5. Glycolysis and related pathways affected by a decrease of PK activity.
(**A**) Metabolic pathways affected due to a decrease in PK activity. Red arrows: metabolites increased on PKD. Dotted red lines: allosteric inhibition caused by increased metabolites. (**B**) Overview of metabolic routes affected. Green arrow: via augmented due to PEP accumulation. Red arrows: inhibitory effect increased by 2, 3 BPG increment. Red lines: Inhibited pathways due to lack of reductive power.

RBC do not have nuclei nor mitochondria, so decrease in PK activity causes impairment of the main source of energy of the erythrocyte. All the processes that require ATP would be affected.

A good example is the activity of the "floppases", that maintain the asymmetric distribution of lipids in the membrane in an energy-dependent fashion[42]. Other affected functions are Na^+-K^+-ATPase and Ca^{2+}-ATPase (plasma membrane Ca^{2+}-ATPase, PMCA) that preserve high K^+ and low Ca^{2+} concentrations inside RBC. Consequently, ATP deficit turns out in important osmotic alterations.[43,44]

Additionally to the mayor features above described, lack of PK activity also affects other important metabolic processes in this cell type:

Methemoglobin (mHb) increase. Apart from ATP, one of the most important products of glycolysis is the reduced form of nicotinamide adenine dinucleotide (NADH). NADH is a cofactor crucial for the activity of NADH-cytochrome b5 reductase (cytb5r), the main enzyme capable to reduce methemoglobin to functional hemoglobin. A deficiency in RPK activity decreases the source of NADH, and consequently, impairs RBC capacity to reduce mHb.[36,45,46]

Glutathione (GSH) reduction. GSH is the main antioxidant agent present in RBC. The presence of GSH is necessary to avoid the toxic effect of oxidant compounds. The thiol groups of GSH are capable of reducing hydrogen peroxide (H_2O_2) to water (H_2O) under the action of the GSH peroxidase. To be able to carry out this function, glutathione must be in its reduced state. GSH is reduced from glutathione disulphide (GSSG) thanks to GSH reductase, which requires nicotinamide adenine dinucleotide phosphate (NADPH). The main source of NADPH comes directly from the second step of glycolysis, where glucose-6-phosphate enters in the hexose monophosphate shunt (also called pentose phosphate pathway). Some authors point out that NADPH generation can also be disrupted by the increased 2,3-BPG. This compound competes with glucose-6-phosphate and 6-phosphogluconate, substrates of two reactions of the pentose phosphate pathway, stopping the NADPH production.[40]

Increased O_2 release. Another erythrocyte function that changes because of the alteration of glycolysis is related with the affinity of Hb-O_2 binding. It has been described

that the increase of organic phosphates, such as 2, 3-BPG, in RBC produces a decrease of the oxygen affinity. In this case, this lower affinity acts as a compensatory mechanism of the anemia since oxygen is more easily released to tissues.[47]

As a result of all these alterations, PK deficient erythrocytes are weaker than the normal ones, shortening their lifespan.

Pyruvate kinase deficiency

1. The disease and palliative treatments

As mentioned above, PKD is produced by mutations in the *PKLR* gene. PKD is considered a chronic hemolytic anemia. Indeed, it is the main cause of chronic non-spherocytic hemolytic anemia (CNSHA) and hence, patients have to deal with it all over their lives. Despite the disease tends to stabilize during adulthood, exacerbations may occur due to different factors, such as pregnancy or infections.[48]

There is not consensus about the prevalence of PKD, some authors estimate 5:10,000 whereas others consider 1:20,000.[48–50] In any case, both estimations are low prevalence values, so it is considered a rare disease.

About a 80% of PKD patients need blood transfusions, half of them just occasional transfusions and the other half are transfusions dependent.[51] Their Hb values vary from 11 g/dL in mild cases to <8.5 g/dL in the most severe cases (while healthy values are 12-17 g/dL). The most severe cases are the ones that are transfusion dependent even after splenectomy.[35,48,49]

Another classical feature is the increased reticulocyte percentage that egress to the blood stream. Bone marrow tries to compensate anemia producing more erythrocytes, thus, reticulocyte percentage exceed the normal 2% threshold, reaching up to 30% in some severe PKD cases.[35,49,52]

Apart from misdiagnosis of mild deficiencies or even non-diagnosed cases, there is a cohort of PKD cases hard to detect. These are the severe cases that affect the unborn patients which lately turn in intra-uterine death.[48] It happens that PKD is especially dangerous in perinatal stages because newborns are not able to compensate anemia. Hence,

PKD is associated with hydrops fetalis, a generalized edema that presents with cardiac failure, threatening fetus´s life.[53]

Furthermore, the direct consequence of hemolytic anemia after birth is neonatal jaundice. This condition is present in 30-50% of the cases if the anemia is mild or moderate and up to 90% if it is severe. Treatment varies from phototherapy (93% of the cases) to exchange transfusion (46% of the cases) depending on severity.[48,51]

As it has been explained before, PKD erythrocytes are fragile and tend to suffer hemolysis. Similarly as described in other hemolytic anemias, PKD produces increase in bilirubin level, a sub-product of Hb metabolism, and decrease haptoglobin level, the protein which binds free Hb to avoid its oxidative action.[50]

Equally important is the loss of RBC that are retained and eliminated in the spleen. In normal conditions, spleen is the agent which removes senescent RBC from the circulation. Since PKD erythrocytes are impaired spleen detects them as abnormal and destroys them. Consequently, spleen´s overwork increases organ´s size producing splenomegaly. Besides, spleen eliminates reticulocytes too, which partially counteracts the compensation that bone marrow develops.[54,55] Thus, splenectomy is carried out as palliative treatment for PKD. This strategy is meant to avoid RBC destruction, either reticulocytes or abnormal erythrocytes, increasing the global number of cells transporting oxygen. Splenectomy does not arrest hemolysis but results in an increase of Hb values around 1-3 g/dL and reduce or abolish the transfusion requirement in most cases. Moreover, it also increases reticulocyte percentage reaching 70% in some patients.[35,48,49] However, this medical practice implies several complications, such as risk of sepsis because of defective defense against encapsulated organisms, thromboembolic disease (that affects 11% of splenectomized patients)[51], injury to the tail of the pancreas, hepatic or portal vein thrombosis, among others.[5,48]

One of the collateral problems of these two palliative treatments is iron overload. In few cases, PKD patients show special susceptibility to iron overload caused by mutations in iron metabolism.[56,57] Apart from these, blood transfusion is principal actor involved in iron overload, affecting 48% of splenectomized patients.[51] Serum ferritin values rise after each transfusion and the lack of spleen impairs the recycling capacity of iron species. This

condition requires appropriate medical control, for instance, based on the use of iron chelation. Otherwise, it can be life-threatening.[48,58,59]

There are some others frequent clinical manifestations such as gallstones (that appear in 45% of splenectomized patients and 48% of which require cholecystectomy)[51], leg ulcers, splenic abscess, kernicterus or acute pancreatitis, all of them related with alterations of iron and hemoglobin metabolism.[31,45,58,60,61]

Recently, it has been published another possible approach to minimize PKD. AG-348 is a PK allosteric activator that is able of increasing the enzymatic activity even in some mutated PKs, so it has been proposed as possible treatment. [62] A phase II clinical trial started in 2015 in which 25 patients were treated ≥3 weeks with AG-348. After treatment, 13 PKD patients (52%) showed hemoglobin (Hb) increase >1.0 g/dL. Differential responses were probably due to variability on mutations since AG-348 depends on mutated PK structure to work.[63]

2. *PKLR* mutations

PKD is an autosomal recessive disorder and most patients are compound heterozygous. Up to 256 variants of *PKLR* gene have been sequenced, among which only 6 are polymorphisms that do not affect the protein sequence (Figure 6, A). Among exonic mutations, single substitution is the most common type of mutation, 72% (67% missense and 5% nonsense), while deletion, insertion or deletion plus insertion are minority (9%, 3% and 1%, respectively). [50]

Apart from mutations in exons, it is demonstrated that PKD can be produced because of alterations in promoter regions, such as GATA-1 binding site, or because of changes of splicing sites within intron regions.[50]

In general, mutations are spread all over the gene being more frequent between exons 7 and 11 (Figure 6, B). Some mutations are more reported in a geographic region than others. For instance, 1529G>A is the most common mutation in north Europe and USA (42%), 1456C>T in south Europe (41%) and 1468C>T in Asia (29%).[35,48]

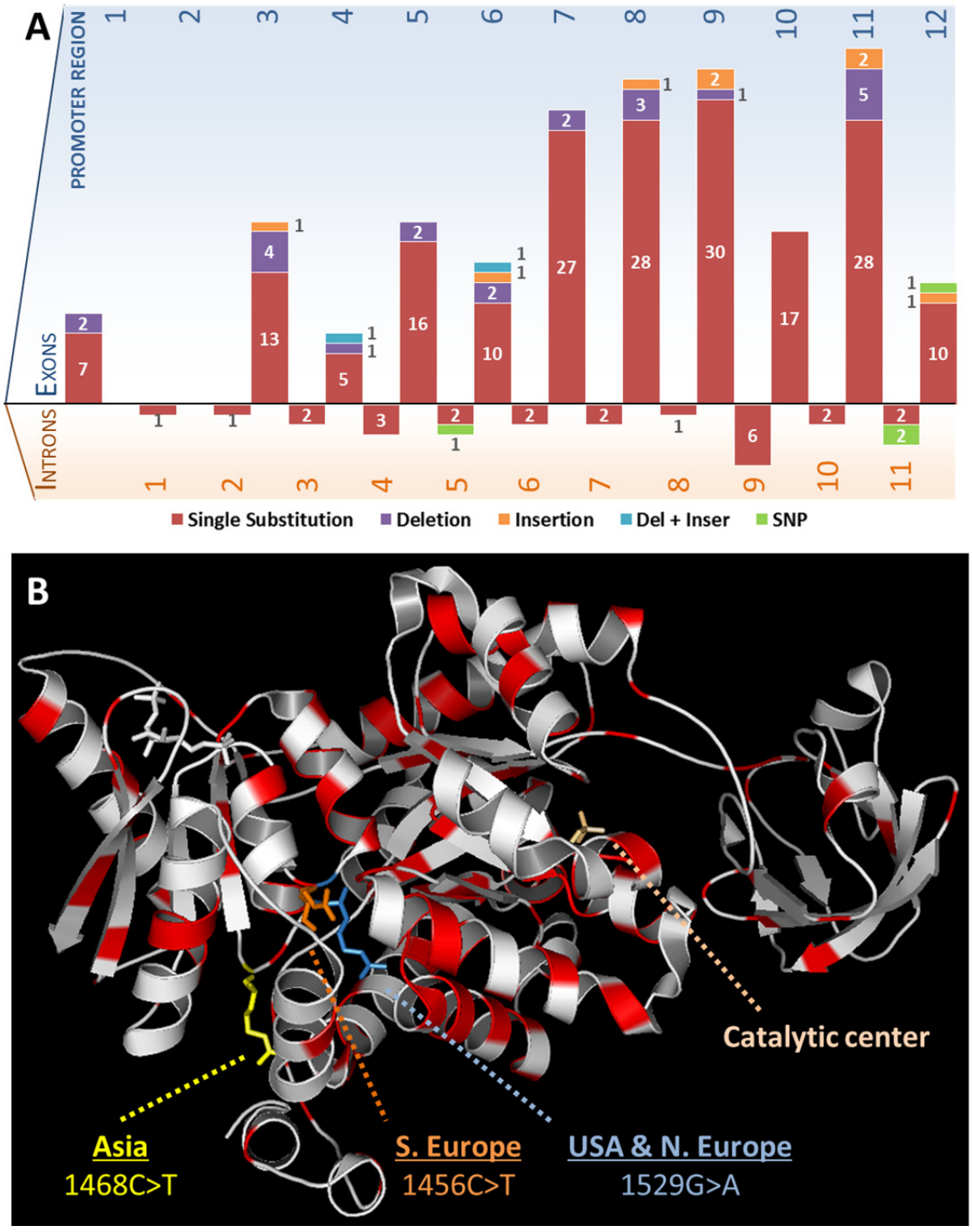


Figure 6. Reported mutations along *PKLR* gene and RPK protein.
(**A**) Mutations causing PKD reported all over exons (top) and introns (bottom), including promoter region. Red: single nucleotide substitution; purple: deletions; orange: insertions; blue: deletions plus insertions; green: SNPs that do not affect the protein sequence. Numbers inside the bars are the amount of different variants found. (**B**) PK monomer showing, in red, PKD mutations described. Yellow, orange and blue are the most represented mutations in different geographic regions. Model based on PDB: 2VGB [18]

The official reports claim that more than 80% reported cases are placed in Europe and Asia (61% Europe, 22% Asia). But probably it doesn't represent the real distribution of the disease because it barely includes data from Latin American or African territories. It is also known that PKD confers protection against malaria; consequently, a higher incidence would be expected in places where malaria is present. Africa, for instance, represents less than 5% of the mutations described.[48,64,65]

Finally, some ethnic minorities present an abnormally high mutation frequency. This phenomenon is justified by founder effect and endogamy. These are the cases of the Ohio and Pennsylvania Amish (1436G>A) and the gipsy communities (deletion of exon 11).[50,66,67]

3. Diagnosis

Some of the symptoms described in the previous section can be useful to diagnose PKD. Reticulocytosis, splenomegaly, increased bilirubin, are normal features among hemolytic anemias, so the difficulty lies in distinguishing PKD from other hemolytic anemias. Classically, PKD has been related with the presence of spiculated erythrocytes. This phenomenon, named echinocyte, correlates with a decrease of K^+ level as described for PKD.[5] Nevertheless, it cannot be used as a differentiating factor because in numerous PKD cases erythrocytes have shown normal shape.[35] Furthermore, echinocytes are also present in other anemias and disorders, such as renal failure.[5]

Thus, the unique biochemical marker characteristic of PKD is the decrease in PK activity. Normally, it is expressed in UI/gHb but it can be found as pyruvate kinase activity/hexokinase activity ratio (PK/HK). Reticulocytes have higher PK activity than erythrocytes. Hence, to make PK activity reliable, this augmented activity in reticulocytes is normalized by HK, which also decreases during RBC maturation. PK/HK ratios between 7.2-15.6 are considered normal. Values under this threshold can be considered PK deficiency markers.

Regarding PK activity evaluation, some authors have suggested that PKD can be compensated for a persistent presence of PKM2 in mature erythrocytes.[68,69] On the other hand, most authors claim that PKM2 activity cannot be detected in mature PKD erythrocytes. Furthermore, it seems that PKM2 activity detected in some cases must be due to an incomplete removal of leucocytes during purification.[48,50,70]

Although the diagnosis of PK deficiency is based on the demonstration of a reduced PK activity, the final diagnosis must be based on a complete genetic sequencing of *PKLR*. Normally, the study is also performed in the parents in order to trace the origin of the mutations.

After diagnosis, disease is classified based on its severity. The guides for this classification are hemoglobin level, frequency of blood transfusions and the need of splenectomy. Also, the improvement of these parameters after splenectomy is taken into account. Normally, patients are divided in 3 groups: mild, moderate and severe. Sometimes a fourth group, very severe, is included, for those who are transfusion-dependent even after splenectomy (Table 1).[48,49]

Table 1. Patient classification based on the severity of their phenotype.

		I. Mild	II. Moderate	III. Severe	IV. Very Severe
Neonatal features	**Jaundice**	30-50% cases	30-50% cases	90% cases	90% cases
	Transfusions	Rare	Rare	Always	Always
		I. Mild	**II. Moderate**	**III. Severe**	**IV. Very Severe**
Adult features	**Splenectomy**	Rare	Sometimes	Always	Always
	Pre- splenectomy				
	Hb (g/dL)	>10	8-10	<8	<8
	Transfusions	Rare	Common	Periodic	Periodic
	Post- splenectomy				
	Hb (g/dL)		10	≥8	<8
	Transfusions		Only exacerbations	Only exacerbations	Periodic

Interestingly, genotype/phenotype association is not clear. Sometimes specific mutations have been associated with a tendency to produce concrete phenotypes (e.g. homozygous for 1456C>T tends to produce mild PKD cases).[48] But the study of 27 patients homozygous for the same mutation (1529G>A) showed that multiple outcomes are possible having the same genetic profile. Among these 27 patients, Hb values vary between 4.9 and 12.2 g/dL and their reticulocyte percentage from 4 to 66 %.[35] However, other authors have reported a worse prognosis in the case of patient with double non-missense mutations (including nonsense, in frame small insertion/deletion, large deletions and mutations in splicing sites) compared with double missense (M/M) or compound missense/non-missense (M/NM) genotype.[35,51]

Curative treatments for PKD

1. Hematopoietic allogeneic stem cell transplantation

As others hematopoietic genetic diseases, the only way to completely cure PKD is to reproduce a healthy hematopoiesis. Up to the date, the unique curative treatment carried out in humans is hematopoietic allogeneic stem cell transplantation (HSCT). Until 2017, only 4 cases of curative outcome had been reported despite of the fact that there were more PKD patients transplanted.[59,71,72] A recent study has collected information of a total of 16 PKD transplanted patients. The life-threatening drawbacks of HSCT for PKD patients are the common in this kind of transplant: graft versus host disease (GVHD) and infections. Survival rate after a 3 year follow-up was 65% (7/16 have not reach 3 years milestone yet). Transplant outcome was clearly better in the case of patients younger than 10 years old (90%).[73] This value is lower compared to the survival percentage of most HSCT for different alterations, such as malignancies or autoimmune diseases (73-93%)[74] but higher than in HSCT for other anemias.[75]

One of the most harmful complications, GVHD, appeared in 38% PKD patients. In this case, it is higher than in other allogenic transplant for other inherited hematopoietic diseases, such as Fanconi anemia (<19%), but lower than Diamond Blackfan anemia (57%).[74–76]

This important drawback of HSCT, caused by haplotype discrepancies, resulting in immune responses that threat the treatment success could be overcome by means of transplant of genetically corrected autologous hematopoietic progenitors.

2. Gene therapy, addition strategy

Monogenic diseases, like PKD, are the perfect scenario for gene therapy. In this kind of diseases, the most common strategy consists in the addition of functional copies of the gene that is mutated in the patients. Corrected patient cells are able to recover the protein/function that causes the disease. There are multiple ways to deliver the DNA into patient cells, among which the use of viral vectors is the most common. Numerous gene therapy clinical trials based on this approach have been performed and are still ongoing.[77] Gene therapy for PKD has not reached that point yet.

Some pre-clinical trials have been carried out showing the feasibility of gene therapy in PKD. Back in the 90´s, hematopoietic progenitor cells from healthy C57BL/6 mice were transduced with a retroviral vector containing the coding DNA of the human *PKLR* and transplanted into lethally irradiated C57BL/6. *PKLR* mRNA was detectable in recipient bone marrow until 135 days post-transplant.[78] In our previous work, published in 2009, we used a PKD mouse model (AcB55) to demonstrate the phenotypic rescue after gene therapy. In this case, the vector used was a γ-retrovirus backbone containing the spleen focus-forming virus (SFFV) promoter and the human RPK cDNA. Correction of 25% bone marrow cells was sufficient to rescue the animals from the main clinical features.[79]

In 2012, another PKD animal model was used to test the feasibility of gene therapy to treat PKD. It was a Basenji dog, a race in which this disease is relatively common. PKD in dogs is pretty similar to the human deficiency, being life-threatening in some cases. This dog, tagged H145, was treated by gene therapy using a foamy viral vector. To drive the transgene expression, PGK promoter (phosphoglycerate kinase promoter) was chosen. PGK promoter provides a high constitutive expression and it is commonly used in gene therapy. In this case, the transgene was the canine RPK. As happened in the mouse model, the therapy cured the dog, making it transfusion-independent.[80]

However, gene therapy scenario changed after the report of insertional oncogenesis cases produced by γ-retroviral vectors. It turned out that γ-retroviral vectors tend to integrate near proto-oncogenes, such as cell cycle regulatory genes, with dramatic consequences.[81–86]

A long way towards safe gene therapy has been walked. Many efforts have been focused on developing safe and efficient viral vectors. Among the new vectors used, self-inactivating lentiviral vectors have shown good transduction levels and there are no reported cases of insertional mutagenesis in any of the more than 150 opened clinical trials so far.[87]

Recently, our group has shown PKD correction using a lentiviral-based gene therapy. The construct uses an internal PGK promoter to express a codon optimized version of human RPK cDNA. PKD mice (AcB55) recovered a healthy phenotype after the transplant of corrected hematopoietic progenitors containing about 1 viral copy per cell.[87] The success of these pre-clinical trials allowed us to apply for orphan drug designation, obtaining it from

EMA (European Medicines Agency) and from FDA (Food and Drug Administration) in 2014 and 2016, respectively. A clinical trial is planned to be opened on 2019 to test the developed medical product. Before this, some additional information is required. For example, what minimum viral copy number of therapeutic vector is required to restore a healthy phenotype is fundamental to decide the transduction efficacy required for a therapeutic treatment.

3. Gene editing approaches

One of the characteristics of addition gene therapy is that the abnormal protein is not removed and coexist whit the corrected one. However, sometimes this fact can impair the efficacy of the therapy due to competition between the endogenous and the transgene protein. Moreover, addition gene therapy does not reproduce the natural regulation of the expression of the gene introduced.

Our group has also been working in a gene editing-based approach to correct PKD. Induced pluripotent cells (iPSCs) derived from PKD patient cells were corrected by homologous recombination. In this process, genomic DNA was cut at intron 2 thanks to the activity of specific TALEN. A DNA matrix containing an optimized version of *PKLR* gene was introduced by homologous recombination. After recombination, codon optimized sequence was located after endogenous 2nd intron, therefore, the new chimeric gene expression remained under the original promoter regions. This strategy permitted the transgene expression under the endogenous tissue specific signals and also interrupted the expression of the deficient copy of *PKLR* gene, avoiding isoform competition. Cells were amplified *in vitro* and their energetic deficit due to PKD was corrected.[88]

PKD models and gene editing tools

1. PKD animal models

As it has been briefly commented above, several animal species that suffer pyruvate kinase deficiency have been used as models to study the disease.

Different mutations have been described in canine *PKLR* gene. Some affected dog breeds, such as West Highland white terrier[89–92] and Cairn terrier[93], present a 6 bp insertion in exon 10 that produce deficiency on RPK. Other mutations have been found in different

breeds: in Beagle (missense mutation, c.994G>A, exon 8)[91,93], in Labrador retriever (truncating mutation, c.799C>T, exon 7)[93], in Pug (missense mutation, c.848T>C, exon 7)[93] and a deletion of a C (c.433delC) in Basenji dog.[80,94–96] Among these dog breeds, only Basenji dogs have been used as PKD model for innovative therapies as it has been described in previous section.[80]

PKD has been also shown in domestic cats. Although it is less studied at genetic level, it is stablished that this condition affects different cat breeds, especially in Abyssinian and Somali cats.[97] Nevertheless, a study on different cat strains (Bengals, Egyptian Maus, La Perms, Maine Coon cats, Norwegian Forest cats, Savannahs, Siberians, Singapuras, Abyssinians and Somalis) indicated that a polymorphism (c.693+304G>A) shared by PKD cats among all this types could be maybe the cause of the disease.[98] Although feline PKD model is normally considered in veterinary field, it has not been used to test any innovative therapies such as gene therapy.

Regarding mouse models, it was back in 1995 when Masahiro Morimoto and collaborators realized that an inbred colony of CBA/N mice that they had in SLC Haruno farm (Shuchi-gun, Shizuoka, Japan) presented splenomegaly and nonspherocytic hemolytic anemia. Also, a marked reticulocytosis was also observed (41.6%) compared with healthy CBA/N (2.8%). These features were due to a homozygous mutation in exon 8 of *Pklr* gene (1013 G>A). They called this strain CBA-Pk-1^{slc}/Pk-1^{slc}. Pk-1 was the old name for *Pklr* gene and slc named the farm where they found the mutated strain. This missense mutation produced a change at PEP binding site (338 Gly>Asp) reducing 90% PK activity. This model was used to demonstrate the feasibility of bone marrow transplantation to correct PKD. [99,100]

The other mouse model, the mentioned AcB55, was found during a screening to find susceptibilities against malaria.[101] It was the inbreeding of a malaria-susceptible strain (A/J) and a resistant one (C57BL/6J) what generated a new strain with a mixed genetic background (approximately 87.5% and 12.5%, respectively). These new malaria resistant mice had a mutation in exon 2 (269T>A) of *Pklr* gene that induced an Ile90Asn substitution. AcB55 mice present reticulocytosis and splenomegaly due to the PKD chronic anemia and have been used as model in this work.[65,102]

Despite having several animal models available, there is not a PKD model based on human cells. Moreover, the normal follow-up of the patients does not include bone marrow aspirations so hematopoietic progenitors collection to study the disease is quite limited.

2. Strategies to generate human disease model in human cells

Classically, drugs, such as enzymatic antagonists, have been used to reproduce protein impairment *in vitro* to generate disease models. But the effect stops once the drug is not present. To obtain a more stable human PKD model, two approaches can be followed, decreasing mRNA level of the gene of interest or directly modifying the gene thereby avoiding even the mRNA release.

2.1 Short hairpin RNAs

A first approach to generate stable disease-like cells is the use of lentiviral vectors containing short hairpin RNAs (shRNAs) to knock down the desired mRNA. These small RNAs are compound for two self-complementary sequences separated by a few nucleotides that form a loop. These sequences are also specific to the mRNA of interest. After being synthetized in the nucleus, pre-shRNAs are processed and transported to the cytoplasm. Once in the cytoplasm, Dicer enzyme with Tat–RNA-binding protein (TRBP) and PACT (PKR activating protein) process the loop of the hairpin to transform them into mature double-stranded siRNAs with 2 nucleotides 3′ overhangs. These mature siRNAs are then ready to act, associated with the RNA-interfering silencing complex (RISC) scan the mRNA of interest and eliminate it.[103,104]

This strategy has been extensively used *in vitro* in cell lines for basic research[105,106], *in vivo* in different animal models[107–109] and also to generate human disease-like cells[110,111]. A similar strategy could be used to generate PKD-like human progenitors.

2.2 Gene editing using nucleases

Other strategy used to generate gene deficient cells is the application of the recently developed gene editing tools to induce DNA breaks in the target gene. After these breaks, endogenous cell machinery repair them by non-homologous end joining, an error-prone mechanism that alters the original sequence resulting in gene disruption. Several nucleases have been developed and used to generate these DNA breaks, such as homing endonucleases

(meganucleases), zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR) together with Cas (CRISPR associated) nuclease.[112–115] Among all of them, the latter one has demonstrated being the most versatile. For this, CRISPR/Cas9 technology was the selected one to modify permanently healthy hematopoietic progenitors in this thesis to generate human PKD-like cells.

CRISPR/Cas is part of the adaptive immune system of prokaryotes against viral hazard. Despite of the fact that the number of proteins that form the CRISPR differs among organisms, all of them share the same goal, the elimination of viral DNA to stop their invasion. In nature, CRISPR are formed by arrays of small sequences complementary to viral DNA spaced by repeated motifs. After the processing of these small RNAs, they are ready to scan these specific viral sequences and link to them. Cas nucleases, associated to this immune system, are joined to these complexes formed by CRISPR RNA and viral sequences and cleave them, protecting the prokaryotic organism from the viral actions.[116–119]

Normally, the CRISPR/Cas systems used in molecular biology are the simplest ones, for instance those in which nuclease activity is carried out by a single nuclease instead of a protein complex (Figure 7).

Originally, it was composed by three separated elements: the so called crRNA, CRISPR-spacer array that contained specific sequences (about 20 bp-long each) complementary to viral DNA; tracrRNA (trans-activating crRNA), an element that joins to the crRNA and acts as scaffold for the nuclease union; and the Cas 9 nuclease, which joins to tracrRNA and to the DNA motifs called PAM (protospacer adjacent motif), 5'-NGG-3' in the case of *S. pyogenes.*[120] For the gene editing approach used in this thesis to generate human PKD-like cells, crRNA sequences used were specific for the interest gene and crRNA and tracrRNA were used in the artificial linked form as a single sequence called single guide RNA or just gRNA.[122] Thus, gRNA forms a heteroduplex with the specific DNA sequence and the two catalytic domains of Cas9 (RuvC and HNH), generating a double strand break (DSB) between the 3rd and 4th nucleotide at 5' of the PAM region.[116,117,120]

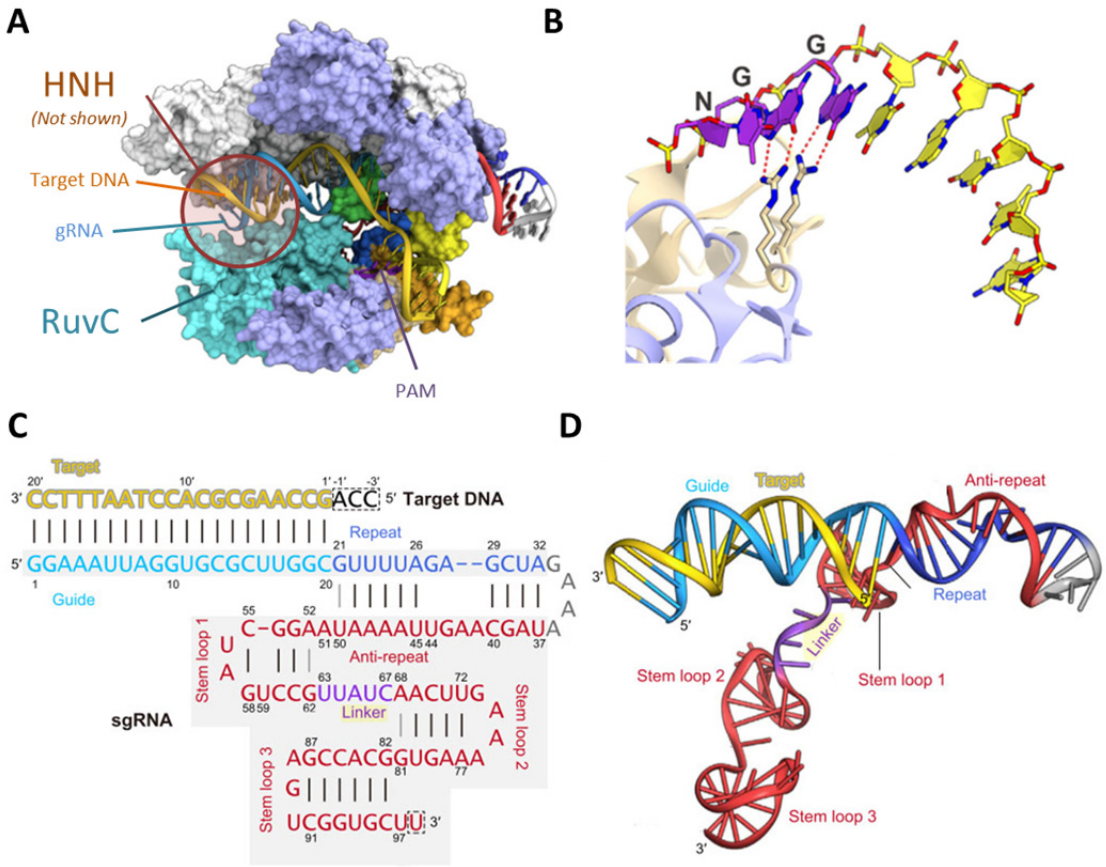


Figure 7 . *Streptococcus pyogenes* Cas9 structure joined to gRNA and targeted DNA.
(**A**) Cas9 protein surrounding target DNA. Yellow strand: targeted DNA; light blue strand: guideRNA complementary to targeted sequence; dark red circle: HNH nuclease domain location, this domain is not shown to visualize DNA heteroduplex; light blue domain: RuvC nuclease domain; small purple region: PAM. (**B**) PAM nucleotides joined to Cas 9´s arginine residues. DNA and gRNA union represented by sequence (**C**) and by 3D structure (**D**). C and D schemes show gRNA generated by joining crRNA and tracrRNA. Figures modified from [120,121].

Objectives

Objectives

Chronic hemolytic anemias, such as PKD, alter normal hematopoiesis increasing the egress of immature erythrocytes, being reticulocytosis and splenomegaly some of its main features. The general aim of this work was to reach a better understanding of the anemic processes that happen in PKD patients through a complete analysis of their peripheral blood. Also, to find out if the compensatory mechanisms carried out by bone marrow affect other cell types such as erythroid progenitors and hematopoietic stem cells, as well as immune cells. Expand our knowledge studying patient samples would help us to understand the disease and would approach us to the eventual clinical trial. To make this clinical trial possible, beyond knowing the disease itself, it is also necessary to know the behavior of the disease when is challenged by gene therapy. The therapeutic vector previously developed by our lab (named coRPK in this thesis) demonstrated its efficacy to cure PKD in an animal model and was designated orphan drug by the EMA and FDA. An evaluation of the performance of the viral vector in limiting situations, to define the scenario in which clinical trial would be conducted is needed. Ideally this experiment should be carried out in human cells, but the source of PKD hematopoietic progenitors to test this and other approaches is limited to the few thousands of hematopoietic stem cells present in the peripheral blood samples. To beat this limitation, it is necessary to generate human PKD-like cells out of healthy $CD34^+$ cells.

For all these reasons, the specific objectives of this thesis were:

1. To further study the extent of hematopoietic alterations in PKD patient samples.
2. To find out if anemic situation affects more primitive populations than reticulocytes.
3. To define the minimum number of corrected cells capable to rescue PKD mice from the deficient phenotype.
4. To generate PKD-like cells from healthy hematopoietic progenitors using two different strategies: mRNA knock-down by short hairpin RNAs and genome editing by CRISPR/Cas9 system.

Materials and Methods

Materials and methods

Primary cells sampling

1. Human hematopoietic progenitors ($CD34^+$)

Hematopoietic $CD34^+$ progenitors were obtained from cord blood (provided by Centro de Transfusión de Madrid) or from peripheral blood (provided by associated physicians) based on the following process. First, total blood was separated by density gradient using Ficoll-Paque PLUS (GE healthcare) according to manufacturer´s instructions. Then, mono-nucleated cell layer was collected and washed with PBE (Dulbecco's Phosphate Buffered Saline, Sigma, herein after named PBS; HyClone 2%, GE Healthcare; EDTA 2 mM). After counting the cellularity, $CD34^+$ cells were labeled with anti-CD34 antibody joined to magnetic beads using CD34 MicroBead Kit (Miltenyi Biotec) and purified by positive selection using two columns, named LS and MS (Miltenyi Biotech), sequentially (attached to QuadroMACS separator and OctoMACS separator magnetic supports, respectively, Miltenyi Biotech). $CD34^+$ cell enrichment achieved, analyzed by flow cytometry, was always above 90%.

In the case of PKD peripheral blood samples, mono-nucleated cells were also isolated using Ficoll-Paque PLUS (GE healthcare) but no $CD34^+$ purification step was performed.

Human blood cell counts were measured on a COULTER Ac•T cell counter (Beckman Coulter).

2. Mouse hematopoietic progenitor sampling

A mouse hematopoietic progenitor cell population, hereafter named Lin^-, was used in this thesis. The name corresponds to its immunophenotype since they are negative for all hematopoietic mature lineages including erythroid (Ter119), T lymphocytes (CD3), mature B, T lymphocytes and monocytes (CD45R), myeloid and NK cells (CD11b) and granulocytes (GR1).

Total bone marrow cells (obtained from femur and tibia) were treated with ammonium chloride buffer (NH_4Cl 0.155 M, $KHCO_3$ 0.01 M, EDTA 0.1 mM; 10 minutes of incubation at room temperature) to eliminate mature red blood cells fraction. Cells were washed with PBS, suspended in the same buffer (10^9 cells/mL), incubated with the antibody cocktail gathered on Table 2 for 30 minutes at 4ºC, washed again with PBS containing 2% bovine serum albumin (BSA, Sigma) and 2% penicillin/streptomycin (Gibco) and purified by fluorescence activated cell sorting using an Influx™ cell sorter (BD).

Table 2. Flow cytometry antibodies used to purify mouse hematopoietic progenitors.

	Epitope	Fluorochrome	Clone	Provider	Catalogue #	V_{ab}/100µL (µL)
Lin^-	Ter119	PE	TER119	BD Pharmigen	553673	0.1
	CD45R (B220)	PE	RA3-6B2	BD Pharmigen	553090	0.2
	CD3	PE	145-2C11	BD Pharmigen	553064	0.2
	CD11b	PE	M1/70	BD Pharmigen	553311	0.2
	GR1	PE	RB6-8C5	BD Pharmigen	553128	0.2

3. Mouse peripheral blood sampling

Peripheral blood samples were taken from the tail vein and collected on EDTA coated tubes (Microvette CB 300 K2E, Sarsted). Complete blood count was measured using an abacus Junior Vet (Diatron).

Experimental animals

PKD mouse model used in this thesis was the recombinant congenic mouse AcB55 (Emerillon Therapeutics Inc., Montreal, Quebec, Canada) which genetic background was 12.5% of A/J and 87.5% C57BL/6 (B6). Healthy wild type C57BL/6 (B6) and A/J mice were used as controls. All, PKD and control mice were maintained at the Laboratory Animal Facility of CIEMAT (28079-21A) by inbreeding.

Mice were maintained under standard diet (water and food *ad libitum*) and all protocols were done following the European and Spanish laws and regulations (ETS123 European convention about the use and protection of vertebrate mammals used in experimentation and other scientific purposes and Spanish law RD 53/2013).

Primary cell cultures

1. Human hematopoietic progenitor medium

Human hematopoietic progenitors were cultured in StemSpan SFEM I (Stem Cell Technologies) supplemented with human stem cell factor (hSCF, 100 ng/mL; EuroBioSciences), recombinant human thrombopoeitin (hTPO, 100 ng/mL; EuroBioSciences), human FMS-like tyrosine kinase 3 ligand (hFlt3-ligand, 100 ng/mL; EuroBioSciences), Glutamax (1%; Gibco), penicillin/streptomycin (1%; Gibco). Cells were maintained at $1x10^6$ cell/mL concentration. Incubation conditions were the following: 37ºC, 5% CO_2 and 95% relative humidity.

2. Erythroid differentiation protocol

Human hematopoietic progenitors were differentiated toward erythroid lineage using 4 different media (Figure 8). The first medium, days 1 to 7, was based on StemSpan SFEM I (Stem Cell Technologies) supplemented with hSCF (50 ng/mL; EuroBioSciences), hFlt3-ligand (16.7 ng/mL; EuroBioSciences), bone morphogenetic protein 4 (BMP-4, 6.7 ng/mL; Peprotech), human interleukin 3 (hIL-3, 6.7 ng/mL; EuroBioSciences), human interleukin 11 (hIL-11, 6.7 ng/mL; EuroBioSciences) and human erythropoietin (h-EPO, 1.3 U/mL; Amgen).

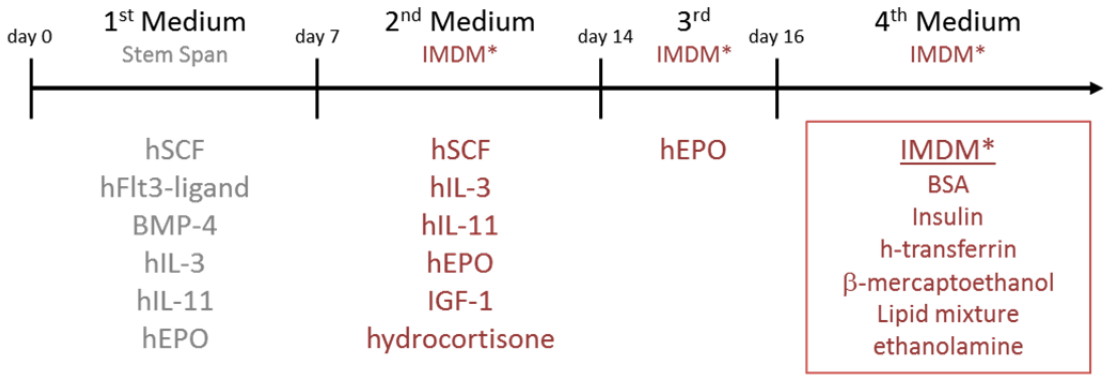


Figure 8. Erythroid differentiation protocol.

Cells were then transferred to the second medium in which they were culture on days 7 to 14. Second medium was based on Iscove's modified Dulbecco's medium Glutamine supplemented (IMDM GlutaMAX supplemented; Gibco) enriched with BSA (1%; Sigma),

insulin (0.01 mg/mL; Sigma), human transferrin (0.2 mg/mL; Sigma), β-mercaptoethanol (91 μM; Gibco), penicillin/streptomycin (1%; Gibco), lipid mixture 1 (1x; Sigma), ethanolamine (0.004%, Sigma), hSCF (5 ng/mL; EuroBioSciences), hIL-3 (6.7 ng/mL; EuroBioSciences), hIL-11 (6.7 ng/mL; EuroBioSciences), hEPO (1.3 U/mL; Amgen), insulin-like growth factor-1 (IGF-1, 20ng/mL; PeproTech) and hydrocortisone (1 μM; Sigma).

After day 14 and for 2 days, cells were transferred to the third medium that was also based on IMDM GlutaMAX medium (Gibco) supplemented with BSA (1%; Sigma), insulin (0.01 mg/mL; Sigma), human transferrin (0.2 mg/mL; Sigma), β-mercaptoethanol (91 μM; Gibco), penicillin/streptomycin (1%; Gibco), lipid mixture 1 (1x; Sigma), ethanolamine (0.004%; Sigma) and hEPO (10 U/mL; Amgen).

Final enucleation was carried out for five days on the forth medium that was also based on IMDM GlutaMAX medium (Gibco) supplemented with BSA (1%; Sigma), insulin (0.01 mg/mL; Sigma), human transferrin (0.2 mg/mL; Sigma), β-mercaptoethanol (91 μM; Gibco), penicillin/streptomycin (1%; Gibco), lipid mixture 1 (1x; Sigma) and ethanolamine (0.004%; Sigma).

All media were filtered by 0.22 μm before use. Sampling and analysis were done within this third week. During erythroid differentiation, cells were maintained between $4x10^5$ and $4x10^6$ cell/mL concentration. Incubation conditions were also 37ºC, 5% CO_2 and 95% relative humidity.

3. Mouse hematopoietic progenitor medium

Mouse hematopoietic progenitors (Lin^-) were maintained on StemSpan SFEM I (Stem Cell Technologies) supplemented with mSCF (50 ng/mL; EuroBioSciences), hTPO (20 ng/mL; EuroBioSciences), hFlt3-ligand (25 ng/mL; EuroBioSciences), hIL-3 (10 ng/mL; EuroBioSciences), hIL-6 (10 ng/mL; EuroBioSciences) and penicillin/streptomycin (1%; Gibco). Cells were maintained at 10^6 cell/mL. Incubation conditions were the following: 37ºC, 5% CO_2 and 95% relative humidity.

Colony forming unit assay

1. Human hematopoietic progenitors

The presence of hematopoietic progenitors on PKD patient peripheral blood was evaluated according with the following protocol. Mononuclear hematopoietic cells were cultured on methyl cellulose (StemMACS™ HSC-CFU completed with EPO; Miltenyi) at 250,000 cells/mL concentration. Triplicates were performed for each sample using 1 mL for each replica. After 14 days of incubation (37ºC, 5% CO_2, 95% relative humidity), colonies were scored using an inverse phase contrast Nikon ELWD 0.3 microscope.

2. Mouse hematopoietic progenitors

Purified hematopoietic progenitors were cultured on methyl cellulose (Methocult GF M3534; Stemcell) at 1500 cells/mL concentration. Triplicates were performed for each sample using 1 mL for each replica. After 7 days of incubation (37ºC, 5% CO_2, 95% relative humidity), colonies were scored using an inverse phase contrast Nikon ELWD 0.3 microscope.

Cell line cultures

K562 cell line (chronic myelogenous leukemia; ATCC: CCL-243) was cultured in Iscove's modified Dulbecco's medium (IMDM; Gibco), HyClone (10%; GE Healthcare) and penicillin/streptomycin (1%; Gibco). Cells were maintained at $1x10^5$-$1x10^6$ cell/mL.

HT1080 cell line (fibrosarcoma cell line; ATCC: CCL-121) was cultured in Dulbecco's modified Eagle´s medium (DMEM; Gibco), HyClone (10%; GE Healthcare) and penicillin/streptomycin (1%; Gibco). Cells were maintained at $1x10^5$ cell/mL.

HL60 cell line (promyeloblast cell line; ATCC: CCL-240) was cultured in Iscove's modified Dulbecco's medium (IMDM; Gibco), HyClone (10%; GE Healthcare) and penicillin/streptomycin (1%; Gibco). Cells were maintained at $1x10^5$-$1x10^6$ cell/mL.

Lymphoblastic cell line (LCL, generated from B cells from healthy donor and kindly provided by Dr. Paula Río, CIEMAT) was cultured in Roswell park memorial institute medium (RPMI; Invitrogen), HyClone (20%; GE Healthcare), penicillin/streptomycin (1%;

Gibco), β-mercaptoethanol (5 μM; Gibco), sodium pyruvate (1mM; Sigma) and non-essential aminoacids (1x; Lonza). Cells were maintained at $1x10^5$-$1x10^6$ cell/mL.

HEK293T cell line (human embryonic kidney cell line competent to replicate vectors carrying the SV40 T antigen; ATCC: CRL-3219) was cultured in Iscove's modified Dulbecco's medium (IMDM; Gibco), HyClone (10%; GE Healthcare) and penicillin/streptomycin (1%; Gibco). Cells were maintained at $5x10^5$ cell/mL concentration.

Incubation conditions were the same for all cell lines used: 37ºC, 5% CO_2 and 95% relative humidity.

Flow cytometry

1. PKD patients characterization

Immunophenotype of PKD patient and healthy donor samples were analyzed by flow cytometry using five different panels specified on Table 3. For panel 1, that corresponds to the reticulocyte study, 5 μL of non-lysed peripheral blood samples were added to 2 mL of PBS plus acridin orange at specified concentration. After 10-30 minutes of incubation, lectures were done not later than 1 hour after acridin orange addition. For the rest of the panels, antibodies were added at concentrations showed in the table, incubated for 30 minutes at 4ºC and then washed with PBA (PBS with 0.2% sodium azide [Merck]; 1% BSA [Sigma]). For panel 1 and 2, peripheral blood samples were pre-diluted in PBA at 1:400 and 1:10 ratios, respectively. After antibody incubation, a lysis step with ammonium chloride buffer was performed to panels 3, 4 and 5 to remove erythrocytes. DAPI (4', 6-diamidino-2-phenylindole) at final concentration of 1 μg/ml were added to identify dead cells. Events were acquired and recorded with a LSR Fortessa (BD Biosciences) and flow cytometry data were analyzed by FlowJo software (BD Biosciences).

Table 3. Flow cytometry antibodies used to characterize PKD patient.

Panel	Reagent			Provider	Catalogue #	V_{final} (mL)	Concentration
1	Acridin Orange			Molecular Probes	A3568	2	0.025 μg/ml
Panel	**Epitope**	**Fluorochrome**	**Clone**	**Provider**	**Catalogue #**	**V_{final} (μL)**	**V_{ab}/V_{final}**
2	CD71	PE	YDJ.1.2.2	Beckman Coulter	IM2001U	50	0.04
	CD235a	FITC	11E4B-7-6	Beckman Coulter	B49206		0.10
3	CD34	PE	8G12	BD Pharimigen	345802	100	0.05
	CD38	FITC	T16	Immunotech	IM0775		0.05
	CD45	APC	2D1	Biolegend	368512		0.10

4	CD3	APC	UCHT1	BioLegend	300439	100	0.04
	CD4	PE	13B8.2	Beckman Coulter	A07751		0.04
	CD8	FITC	B9.11	Beckman Coulter	A07756		0.04
	CD45	APCCy7	HI30	BioLegend	304014		0.03
5	CD3	APC	UCHT1	BioLegend	300439	100	0.04
	CD14	BV711	M5E2	Biolegend	301837		0.02
	CD15	PE	VIMC6	Miltenyi Biotec	130-091-375		0.04
	CD19	PECy7	SL25C1	eBioscience	25-0198-42		0.03
	CD45	FITC	J33	Beckman Coulter	A07782		0.06

2. Erythroid differentiation immunophenotype

Immunophenotype of hematopoietic progenitors differentiated toward erythroid lineage was analyzed by flow cytometry using the antibody combination specified on Table 4.

Table 4. Flow cytometry antibodies used to follow *in vitro* differentiation along erythroid lineage.

Epitope	Fluorochrome	Clone	Provider	Catalogue #	V_{ab}/100µL (µL)
CD36	PE	CB38(NL07)	BD Pharmingen	555455	1
CD45	APCCy7	HI30	BioLegend	304014	1
CD71	PECy5	M-A712	BD Pharmingen	551143	1
CD235a	PECy7	HI264	BioLegend	349111	1

Cells were resuspended in a final volume of 100 µL, labeled for 30 minutes at 4ºC and washed with PBA. DAPI (final concentration of 1 µg/ml) were added and events were recorded with a LSR Fortessa (BD Biosciences). Data were analyzed by FlowJo software (BD Biosciences).

Viral vectors and plasmids

1. Therapeutic and control lentiviral vectors, coRPK and EGFP

Therapeutic (coRPK) and a control enhanced green fluorescence protein (EGFP) vectors used in this thesis are shown on Figure 9.

The coRPK transgene is a codon optimized version of the coding DNA of *PKLR* gene that produces RPK enzyme. The rest of the sequences of the vector are shared by therapeutic and control. Both of them were self-inactivating lentiviral vectors based on pCCL lentiviral backbone. Cytomelgalovirus (CMV) promoter drove the expression of the

primary transcripts in the packaging cell line. Transgenes (coRPK or EGFP) expression was directed under the human phosphoglycerate kinase (PGK) promoter.

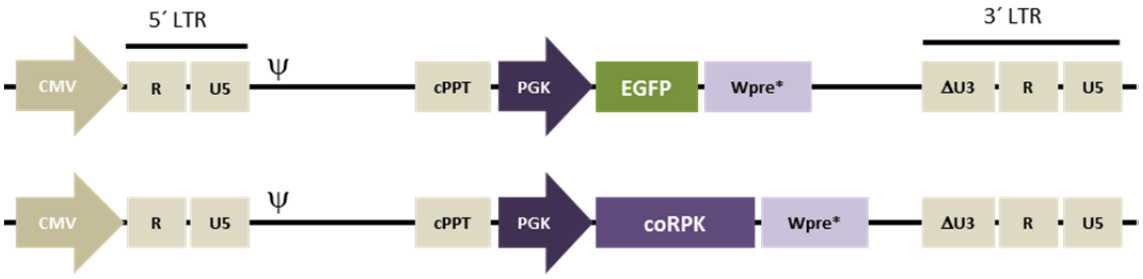


Figure 9. Lentiviral vectors used to transduce hematopoietic progenitor cells of PKD mouse model. Constructs were based on pCCL backbone; expression during viral production was driven by Cytomelgalovirus (CMV) promoter. Transgene promoter was human phosphoglycerate kinase (PGK). Woodchuck hepatitis virus post-transcriptional regulatory element mutated (Wpre*) sequence is meant to stabilize transgene transcript. 3' LTR region was mutated (ΔU3) to make it self-inactivating. EGFP and coRPK constructs only differ in the transgene.

This eukaryotic promoter produces stable expression of the interest genes. A mutated version of Wpre (woodchuck hepatitis virus post-transcriptional regulatory element) was used to stabilize transgene RNA. The mutation included in Wpre sequence prevent the production of a truncated 60 amino acid protein derived from the woodchuck hepatitis virus X gene involved in liver cancer.[123] Viral vectors were packaged using pMD2.VSVg envelope. Both vectors were the same used in preclinical studies carried out in our lab.[124]

2. Short hairpin RNA vectors

Short hairpin lentiviral vectors were based on Wiznerowicz and Trono constructs.[104] This vector contains the short hairpin sequence under H1 promoter expression and also EGFP transgene under human elongation factor-1 alpha (EF1α) promoter (Figure 10, A).

Six different shRNAs were designed against different RPK mRNA regions (Figure 10, B). MluI and ClaI restriction sites flanked the short hairpin sequences to clone it into the mentioned vector. Another restriction site (XhoI) was included into the hairpin loop to corroborate the presence of the specific sequence in the vector after cloning (Figure 10, C). the sequences selected had at least 3 mismatches with respect to coRPK to avoid silencing of the therapeutic transgene (Figure 10, C, red nucleotides).

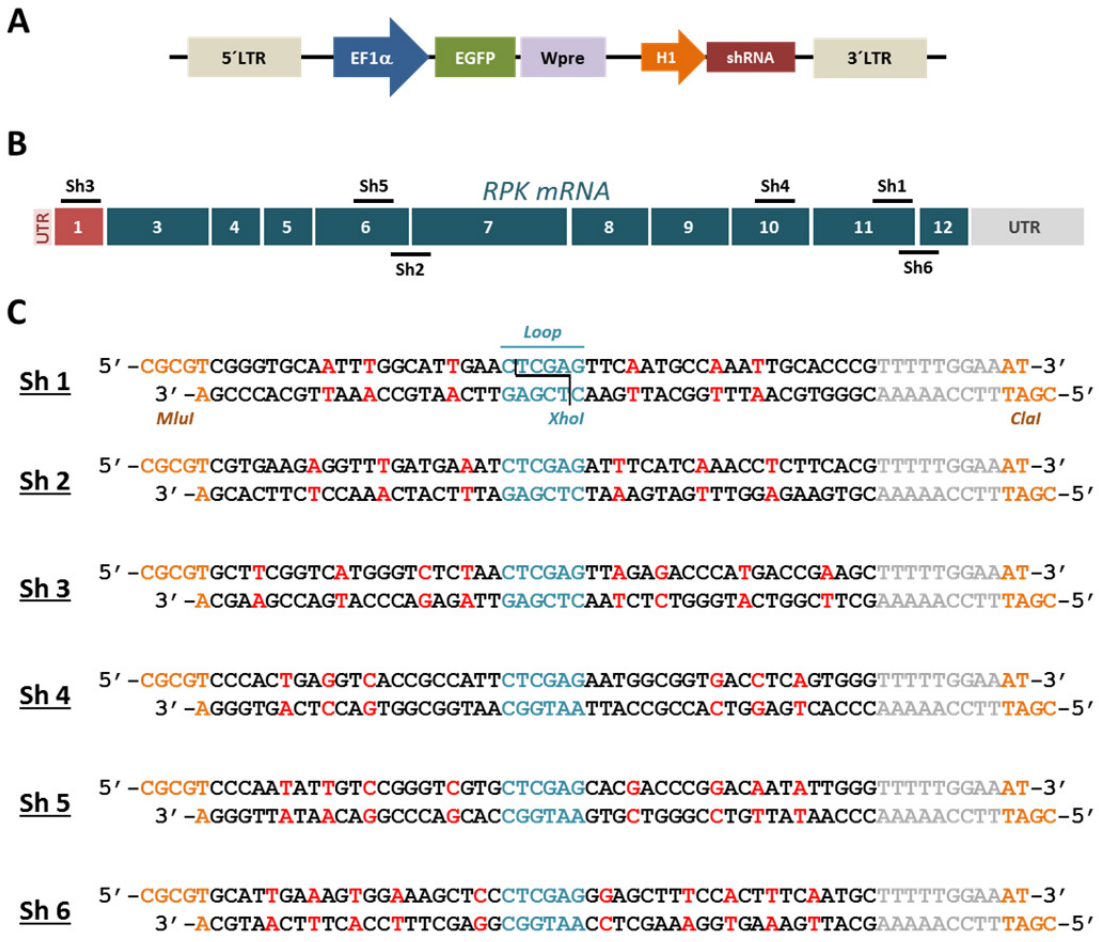


Figure 10. Viral vector, sequences design and location of shRNAs along *PKLR* mRNA.
(**A**) Scheme of the lentivirus used to clone shRNAs (**B**) RPK mRNA divided in exons and the location of the six shRNAs along it. (**C**) shRNA sequences. Orange: cloning restriction sites. Blue: hairpin loop and XhoI restriction site. Red letters: nucleotides that differ with respect to coRPK mRNA.

3. Lentivirus productions and titration

Lentiviruses were produced using a third generation set up with HEK293T line as packaging cells. Productions were carried out in 150 mm dishes (Corning) with 60-70% of cell confluence transferring the different plasmids required by $CaCl_2$ DNA precipitation method.

Plasmids used were kindly provided by Dr. L. Naldini (HSR-TIGET, Milan, Italy) and produced by Plasmid Factory. Plasmids used were: transfer (transgene of interest; 37 µg/p150), *rev* plasmid (pRSV-REV; 9 µg/p150), *gag* and *pol* plasmid (pMDLg/pRRE; 23.5 µg/p150) and VSV envelope (pMD2-VSVG; 12.3 µg/p150). Plasmids were mixed with

457µl/p150 of $CaCl_2$ 2.5 M and 3.2ml/p150 of water. Afterwards, 3.6 ml/p150 of HBS 2x (281 mM NaCl, 100 mM HEPES, 1.5 mM Na_2HPO_4, pH 7) were added dropwise to the mix to generate Ca^{2+}/DNA^- precipitates and this solution was added to the HEK293T cell cultures.

After 5-6 hours, medium was removed and fresh collecting medium was added (IMDM supplemented with 5% HyClone). Supernatant was harvested 48 hours post-transfection, centrifuged (450 g, 7 minutes) and filtered (0.22 µm, PES, Merck) to eliminate cellular rests. Supernatant was concentrated by ultracentrifugation (70000g, 2 hours, 20ºC; Optima L-100 XP Ultracentrifugue, Beckman-Coulter) using Ultra-clear tubes (Beckman-Coulter). Viral pellet was reconstituted on 100-300 µL saline solution (NaCl 0.9%). Viruses were titered by serial dilution of the vector (from 10^{-2} to 10^{-6}) on HEK293T cells (150-180,000 cells/well on multiple well plate size 24; FALCON). In the case of GFP labeled vectors, transducing units (TU) were calculated based on flow cytometry measures 48-72 hours after transduction applying the following formula.

$$\frac{TU}{mL} = \frac{GFP\%/100 \times cell\ number}{V_{viral\ vector}}$$

Where $V_{viral\ vector}$ is the volume of the viral vector stock used for the titration and *cell number* is the number of cells counted in a replicate well at the moment of transduction.

Viral vectors non-GFP labeled were titered with the same dilution protocol and viral copies were measured by qPCR (see *Vector copy number determination*, *Materials and Methods*):

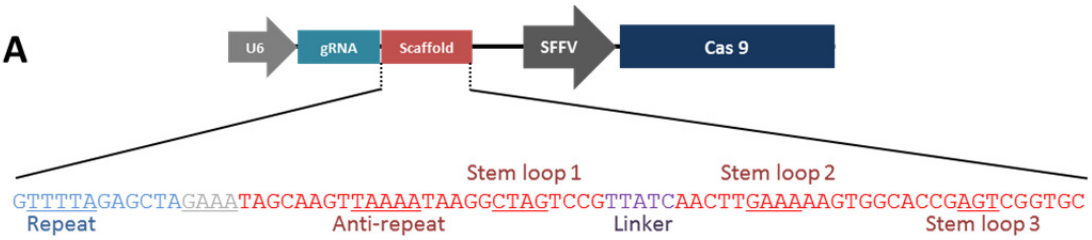
$$\frac{Viral\ particles}{mL} = \frac{VCN \times cell\ number}{V_{viral\ vector}}$$

4. CRISPR/Cas9 plasmids

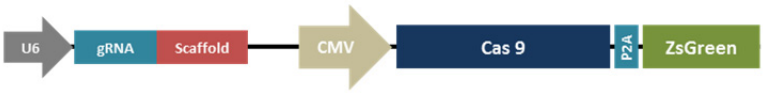
Two different plasmids were used to clone specific guides. Both plasmids had guide RNA and RNA scaffold expression driven under U6 promoter. First construct contained also Cas9 sequence under SFFV promoter (Figure 11, A) and the second one had Cas9 and ZsGreen sequences linked by a P2A (porcine teschovirus-1´s motif that produce co-translational "cleavage" of both proteins) under CMV promoter (Figure 11, B). Both constructs were kindly provided by Dr. Raúl Torres (CNIO, Madrid).

Eleven different guides were designed based on Zhang Lab´s program (MIT)[125] against *PKLR* gene (Figure 11, C). All guide couples named combos had exon 9 guide I in common. Combo´s guides are named after the size of the deletion that generates when are combined with exon 9 guide I (Figure 11, D). Exons 8 and 11's guides were only cloned into the first construct; exon 9's guides were cloned in both constructs and combo 1 to 6 guides were cloned only into the ZsGreen labeled plasmid.

A



B



C

Guide	Score	Sequence	MM respect coRPK	Strand
Exon 8 guide I	79	CGGCATCATGGTGGCACGGGGGG	4	Sense
Exon 8 guide II	72	GAAATCCTGGAGGTGAGCGACGG	4	Sense
Exon 9 guide I	89	TTACCAAGCCCCGGCCAACGAGG	6	Sense
Exon 9 guide II	64	AGCCCCATCCAGCACAGCATTGG	3	Antisense
Exon 11 guide I	95	TCTGTCTCGGTACCGACCTCGGG	8	Sense
Exon 11 guide II	92	CGCCAGGTCCACTTATGCCGAGG	5	Sense
Guide 8	66	TGCCCTCGTTGGCCGGGGCTTGG	6	Antisense
Guide 63	59	CTGACTGCATCATGCTGTCAGGG	2	Sense
Guide 119	79	GTGAAGATGCAGCATGCGGTAGG	5	Sense
Guide 239	78	CAGCTGCCGGTGGTACACTGCGG	1	Antisense
Guide 490	52	GCCTCTCCTCTTGTCTCAGGTGG	7	Antisense

D

Guide pair name	Guide at 5´		Guide at 3´
Combo 1	Guide 8	+	Exon 9 guide I
Combo 2	Exon 9 guide I	+	Exon 9 guide II
Combo 3	Exon 9 guide I	+	Guide 63
Combo 4	Exon 9 guide I	+	Guide 119
Combo 5	Exon 9 guide I	+	Guide 239
Combo 6	Exon 9 guide I	+	Guide 490

Figure 11. CRISPR/Cas9 plasmids and guide RNAs design.
(**A**) Scheme of no-labeled CRISPR plasmid, including guide RNA and Cas9. gRNA: sequences specific to each gene region. Scaffold: DNA sequence that generates the joined gRNA that includes crRNA and tracrRNA joined by the linker represented in purple. Scheme includes structure description of the loops that generate the 3D folding of the complex (see *Strategies to generate human disease model, Introduction*). (**B**) GFP labeled plasmid. (**C**) Guide RNAs including MIT program score, PAM sequence (red) mismatches (MM) with respect to coRPK and the strand which guide is complementary to. (**D**) Summary of the guides that compose each pair of combinations for deletion strategy.

Genetic delivery tools

1. Hematopoietic progenitor transduction protocol

Mouse hematopoietic progenitors (Lin^-) were seeded on 6-well plate (FALCON) previously treated with RetroNectine (20 μg/ml covering the bottom of the dishes overnight; r-fibronectin CH 296, Takara) and pre-stimulated for 24 hours (using medium specified in *Mouse hematopoietic progenitor medium* section, *Materials and Methods*). Then, cells were transduced with the appropriate amount of lentiviral vectors for 24h at final volume of 3 ml/condition. Afterwards, cells were washed with PBS and prepared to be transplanted, or grown depending on experimental design. In the experiment that included a second transduction round, cells were washed prior the second 24 hours transduction.

Human $CD34^+$ progenitor cells obtained from cord blood were amplified for 24-96 hours depending on the experiment using the human hematopoietic progenitor medium previously specified (see *Human hematopoietic progenitor medium section, Materials and Methods)*. Transductions were performed in 96-well plate (FALCON) at final volume of 200 μL. After 24 hours of transduction, cells were washed with PBS and diluted in the first erythroid differentiation medium, also described above (see *Erythroid differentiation protocol section*, *Materials and Methods*).

2. Electroporation of primary cells

Hematopoietic progenitor cells were electroporated with CRISPR/Cas9 plasmids previously described using Amaxa Nucleofector I (Lonza) and following manufacturer´s protocols. Cells were electroporated in cuvettes (electroporation volume of 100 μL) using Human $CD34^+$ Cells kit (Lonza) and U08 electroporation program. Total amount of plasmid used was 11.26 μg, hence, when two plasmids were electroporated at the same time, 5.63 μg of each were added to the mix. After electroporation, cells were cultured on 24-well plate (FALCON) at 2 mL of final volume. In the case of the experiments with selection based on GFP reporter, sorting was performed 24 hours after electroporation. Aggregates were eliminated before sorting using 40 μm Cell Strainer (FALCON).

3. Electroporation of cell lines

CRISPR/Cas9 plasmids were also used in cell lines. DNA delivery was carried out on 20 µL volume format using Amaxa 4D-Nucleofector System and Cell line nucleofector solution SF (Lonza). Total amount of DNA electroporated was 3 µg (1.5 µg each plasmid in the case of using two plasmids at the same time). $2x10^5$ cells were electroporated following manufacturer´s protocol. Different programs were used for the different cell lines (Table 5).

Table 5. Electroporation programs depending on cell type used.

Cell Line	Electroporation program
K562	FF-120
LCL	EW-113
HEK293T	CM-130

DNA purification and analysis

1. Analysis of *PKLR* gene

Corroboration of gene edit events generated by CRISPR/Cas9 system was carried out on PCR amplified DNA. NucleoSpin Tissue kit (Macherey-Nagel) was used to purify genomic DNA. Different primer couples were used to amplify edited regions of *PKLR* exons 8, 9 and 11. Also, a different primer couple, with forward primer at exon 9 and reverse primer at intron 10, was used to amplify the edited region containing the 459 bp deletion generated (Table 6). All primers were designed with NCBI/ Primer-BLAST software and purchased to Sigma.

Table 6. Primers used to amplify gene edited regions.

F= forward primer; R= reverse primer

Exon	F/R	Sequence (5´ to 3´)	Amplicon size (bp)
8	F	GAAGGCGTGAAGAGGTGAGG	300
	R	AACCCACAGAGTGCCGAAC	
9	F	CAGGGGTTGTGACTGTGACC	295
	R	ATTCTGAGCTCCTACCGCAT	
11 guide I	F	CCTCGTTCACCACTTTCTTGC	259
	R	GCAGGGAAGGTCTAGGTAGC	
11 guide II	F	ACTTTCTTGCTGTTCTGGGCT	248
	R	GCAGGGAAGGTCTAGGTAGC	
E9-i10	F	ACAGTGTGAGTCCTACAACTTTGA	670
	R	CAATGTGCCCAGTGATGCCG	

PCR conditions used were the following: genomic DNA (200 ng), dNTPs (0.2 mM), $MgCl_2$ (1.5 mM), forward primer (0.2 µM), reverse primer (0.2 µM), TaqGold DNA

polymerase (1.25 U) diluted in the specific buffer provided by the manufacturer up to a final volume of 50 µL. All reagents were purchased to Applied Biosystems (Thermo-Fisher). PCR program was: 95ºC 10 minutes to activate the polymerase, 40 cycles of 95ºC 45 seconds, 60ºC 45 seconds and 72ºC 2 minutes. Finally, a 10 minutes step at 72ºC for final elongation. Agarose (2%; Conda) gels were used to perform electrophoresis of that PCR amplicons. Ethidium bromide added to the agarose was used to visualize the DNA using GelDoc XR+ Image system (BioRad)

2. Gene editing evaluation

Gene editing evaluation was performed by Surveyor assay on the PCR amplicons. Between 200 and 400 ng of PCR product were used to generate the hetero-duplex composed by wt and edited strands (warming the sample up to 95ºC and letting it cooling 0.3ºC/second). This new DNA mix was incubated with the surveyor enzyme (IDT) for one hour at 42ºC. This enzyme detects single-strand loops generated because of the different sizes of edited and wt strands. Novex™ TBE Gels, 10% (Invitrogen) were used to visualize edited PCR amplicons.

Precise deletions generated using guides I and II for exon 9 were also detected digesting exon 9 PCR amplicon with BstXI restriction enzyme (37ºC, 1 hour; New England BoiLabs). Results were also visualized in 2% agarose gels.

TOPO TA Cloning (Thermo-Fisher) was used, according to manufacturer´s protocol, to clone PCR-amplified DNA fragments of exon 9 after gene editing. These fragments were then sequenced by STABVIDA Company using Sanger sequencing.

Next generation sequencing (NGS) was carried out by STABVIDA Company. PCR products were purified using PCR Clean-up magnetic beads (Axygene) according to manufacturer´s protocol. The samples were used for library construction using the KAPA DNA Library preparation kit for Illumina with KAPA HiFi. The generated DNA fragments (DNA libraries) were sequenced with v3 chemistry in the lllumina MiSeq platform, using 300 bp paired-end sequence reads. Only variants represented in more than a 1% were considered.

3. Vector copy number determination

Vector copy number (VCN) was evaluated on peripheral blood cells of mice transplanted with hematopoietic progenitors transduced with coRPK or EGFP vectors. DNA was extracted from nucleated peripheral blood cells using NucleoSpin Tissue kit (Macherey-Nagel).

VCN was also evaluated on individual methyl cellulose colonies. Each colony was collected individually, centrifuged and stored at -20ºC. Pellets were lysed with 10 µL of PBS and 20 µL of lysis buffer (0.3 mM Tris HCl, pH 7.5; 0.6 mM $CaCl_2$; 1.5% Glycerol; 0.675% Tween-20; and 0.3 mg/mL proteinase K) and incubated at 65ºC for 30 minutes, 90ºC for 10 minutes more and then 4ºC to finish the lysis.

Quantitative PCR used to calculate VCN was based on the simultaneous amplification of ψ viral region of the transgene and mouse´s titin gene as reference (named *Ttn*). Primers and probes used are showed on Table 7.

Table 7. Primers and probes used to perform VCN qPCRs.

F= forward primer; R= reverse primer; P= probe

Amplicon		Sequence (5´to 3´)	Fluorochrome	Quencher
ψ	F	CAGGACTCGGCTTGCTGAAG		
	R	TCCCCCGCTTAATACTGACG		
	P	CGCACGGCAAGAGGCGAGG	6FAM	BHQ1
Titin	F	AAAACGAGCAGTGACGTGAGC		
	R	TTCAGTCATGCTGCTAGCGC		
	P	TGCACGGAAGCGTCTCGTCTCAGTC	Texas Red	BHQ2

Amplifications were performed on a 7500 fast real-time PCR system (Applied Biosystems) using ABsolute qPCR Low ROX Mix (Thermo Scientific).

A double absolute quantification was interpolated on a standard curve generated by serial dilutions (1:10 each) of a plasmid that contained templates for both amplicons (pRRL.hPGKmGFP.Wpre.Mex5.Titin, kindly provided by Genethon). VCN was calculated with the following formula:

$$VCN = \frac{2 \times \psi copies}{Titin\ copies}$$

Cytospin preparation

For cytospin preparations, 100,000 cells were diluted in PBS (containing 20% of HyClon; GE Healthcare) and centrifuged (25 g, 4 minutes) on Polysine™Microscope Adhesion Slides (Thermo-Fisher) on a Cytospin 2 centrifuge (Shandon). Samples were stained with Wright Stain Modified solution (SIGMA) for 3 minutes. Microscope slides were washed with water to remove the excess of staining. Olympus BX41 microscope was used to take pictures and ImageJ (National Institutes of Health) software to score them.

Pyruvate kinase activity/hexokinase activity assay

Two different sample processing were carried out depending on the source of the cells. Mouse red blood cells were obtained taking 5 µL from the bottom layer resultant of centrifugation at low velocity (450 g, 10 minutes). These 5 µL were diluted 1/20 in stabilizing solution (EDTA 2.7 mM, β-mercaptoethanol 0.005%). Cell culture pellets were directly diluted on stabilizing solution. Independently of sample type, after diluting in stabilizing solution, samples were put in liquid nitrogen to fast lysate the cells and stored at -80ºC.

Enzymatic activity assay protocols used were modified from the assay already described[126]. Pyruvate kinase/hexokinase (PK/HK) activity assay is based on two independent enzymatic reactions, PK and HK activities. As long as erythroblasts mature, both PK and HK activities decay. Thus, PK/HK ratio corrects the excess of PK activity generated in the most immature erythroblasts. Each activity was measured based on the following reactions.

Pyruvate kinase activity

$$PEP + ADP \xrightarrow{\boldsymbol{PK}} Pyruvate + ATP$$

$$Pyruvate + \boldsymbol{NADH} \xrightarrow{LDH} Lactate + NAD^{+}$$

Hexokinase activity

$$Glucose + ATP \xrightarrow{\boldsymbol{HK}} G6P + ADP$$

$$G6P + NADP \xrightarrow{G6PD} 6\ Phosphogluconolactone + \boldsymbol{NADPH} + H^+$$

Both assays quantify changes on 340 nm absorbance over time. In PK assay, 340 nm absorbance decreases over time because of the oxidation of NADH to NAD^+, while in HK assay, 340 nm absorbance increases over time due to the generation of NADPH.

For PK assay, reaction mix was prepared on 176 µL (Tris-HCl 0.1 M, EDTA 0.5 mM pH 8, $MgCl_2$ 0.01 M, KCl 0.1 M, LDH 6 U/mL, ADP 1.5 mM and NADH 0.2 mM). Then, mix was incubated with 4 µL of sample at 37ºC for 10 minutes. Finally, 20 µL of PEP (0.1 M) were added to each sample and 340 nm absorbance was measured on spectrophotometer.

Similarly, for HK assay, reaction mix was prepared on 190 µL (Tris-HCl 0.1 M, EDTA 0.5 mM pH8, $MgCl_2$ 0.01 M, NADP 0.2 mM, ATP 10 mM, Glucose 2 mM and G6PDH 0.1 U/mL). Then, mix was incubated with 10 µL of sample at 37ºC for 10 minutes and 340 nm absorbance was measured on spectrophotometer.

Assays were carried out on flat transparent bottom 96-well plate (FALCON) and absorbance at 340 nm was measured for 10 minutes (1 measure each minute) in a GeniousPro reader (Tecan).

Calculations, based on Lambert-Beer law, were also adapted to the low scale assay changing the original 1 cm light path (l) to 0.67 cm. ΔA_{340} (slope of A vs t regression), extinction coefficient of NAD(P)H ($\varepsilon_{NAD(P)H}$) and dilution factor (D) remain like in the original protocol.[126]

$$\frac{U}{min} = \frac{\Delta A_{340}}{min} \times \frac{1}{\varepsilon_{NAD(P)H}} \times \frac{1}{l} \times D$$

Thus, PK and HK slopes must be operated as it is shown in the next formulas:

$$\boldsymbol{PK} \quad \Delta A_{340} \times 1.207 \times (-1000) \qquad \boldsymbol{HK} \quad \Delta A_{340} \times 0.483 \times (1000)$$

Final results were represented as PK/HK ratio.

Protein extraction and western blot

Cells were lysed using lysis buffer (20 mM HEPES, pH 7.5; 100 mM NaCl; 20 mM $MgCl_2$; 10 mM EGTA; 40 mM β-glycerophosphate; 1% Triton X-100; phenylmethane sulfonyl fluoride PMSF 1 mM; protease inhibitor 1x, *Complete Mini Protease Inhibitor Cocktail,* Roche; phosphatase inhibitor 1x, *PhosSTOP*, Roche). Protein content was measured by Bradford method (BioRad Laboratories).

In order to denaturalize and reduce the protein extracts, samples were boiled using a solution 1:9 of β-mercaptoethanol and Laemmli buffer (BioRad), for 5 minutes at 95ºC. Electrophoresis was performed in 4–15% precast polyacrylamide gels (BioRad) on Tris/Glycine/SDS buffer (25 mM Tris, 192 mM Glycine, SDS 0.1%, pH 8.3; BioRad). Protein amount used was 30 µg and Precision Plus Protein Standars Dual Color (BioRad) was used as molecular weight reference. Trans-Blot Turbo Transfer (protocol MIXED MW, 2.5A, 25V, 7 minutes; Biorad) system was used to transfer the proteins to nitrocellulose membranes (BioRad). BioRad transfer buffer with 20% absolute ethanol was used.

Transferred membranes were incubated for 1 hour at room temperature with blocking buffer (PBS containing 5% skimmed milk, 0.1% Tween-20). Primary antibodies, diluted in blocking buffer, were incubated over night at 4 ºC (Table 8). Secondary peroxidase conjugated antibodies were incubated for one hour. After incubation of both antibodies, three PBS washes were performed. Secondary antibodies were detected with Clarity Western ECL (BioRad) following manufacturer´s protocol. ChemiDoc MP System (BioRad) was used to detect the chemiluminescence. Quantity One and Image Lab programs were used to quantify the bands (BioRad).

Table 8. Western Blot antibodies used.

Primary antibodies				
Protein	**Clone**	**Isotype**	**Provider**	**Dilution**
β-actin	AC-15	Mouse IgG	Abcam	1:2,500
γ-tubulin	Polyclonal	Rabbit IgG	Abcam	1:1,000
GADPH	6C5	Rabbit IgG	Santacruz	1:1,000
RPK	Polyclonal	Rabbit IgG	CIEMAT [70]	1:10,000
RPK	Polyclonal	Mouse IgG	CIEMAT [70]	1:3,000
Secondary antibodies				
Protein	**Isotype**	**Conjugated**	**Provider**	**Dilution**
Mouse IgG	Donkey IgG	HRP	Jackson	1:50,000
Rabbit IgG	Donkey IgG	HRP	Amersham	1:10,000

Results

Results

Patients characterization

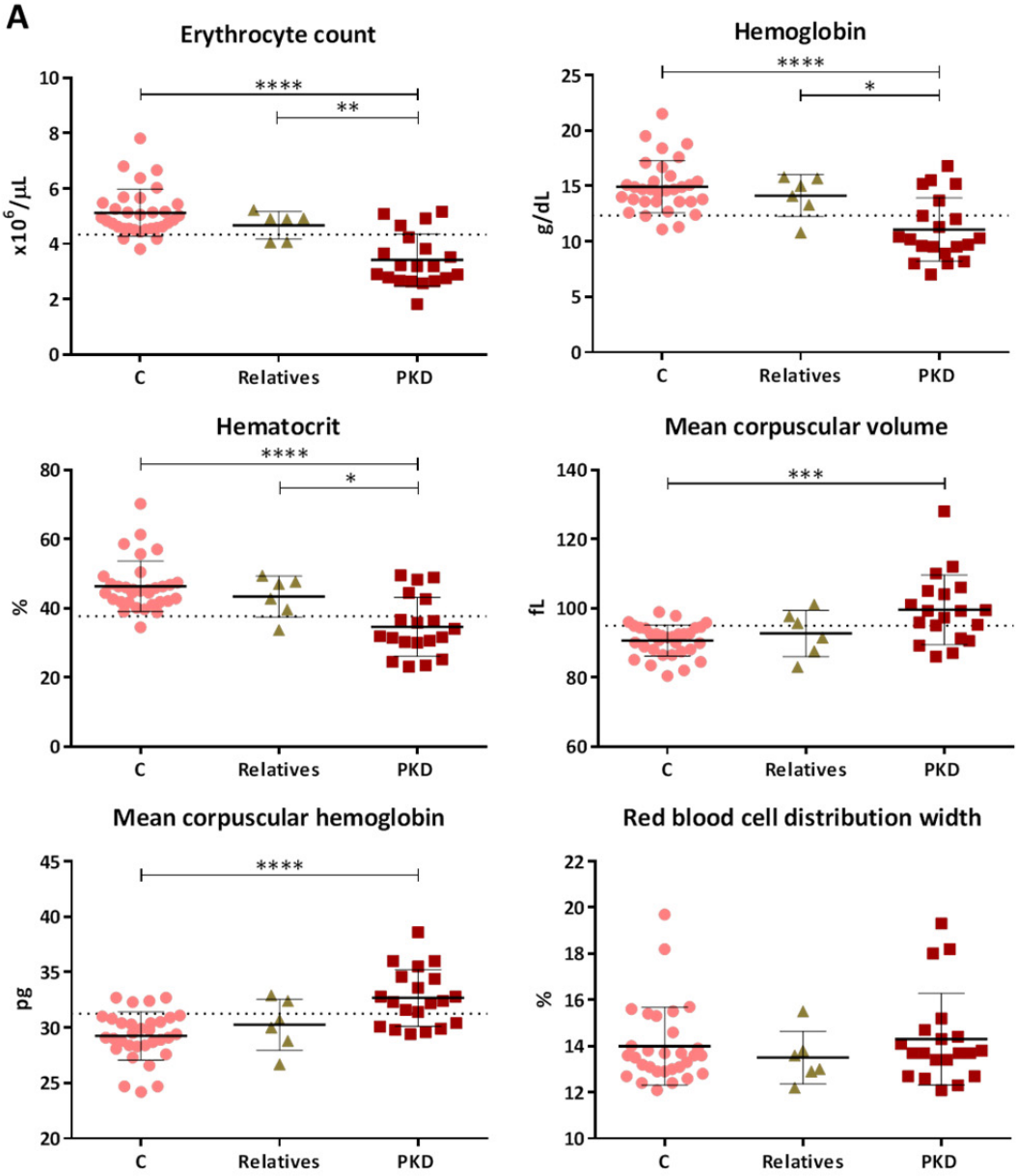
Conventional follow-up of PKD patients is done through periodic peripheral blood sampling and analysis of RBC and white blood cell counts, hemoglobin and reticulocytes among others. More specialized studies, such as immune system populations, late erythroid progenitor profile or circulating hematopoietic progenitor's content, are not performed routinely. All these aspects were addressed here by complete blood count, flow cytometry study of hematopoietic subsets and hematopoietic progenitor quantification by semi solid media culture.

1. Comparative complete blood count

Up to 21 PKD patients and 32 healthy controls were analyzed by complete blood count using an automatic hematological counter. Besides, 6 healthy relatives of PKD patients were also evaluated to study if one mutated allele could produce changes on hematological parameters. Specifically, these relatives were: three mothers, two fathers and one brother from three different patients.

Red blood cell subsets were the most influenced on PKD. The number of RBC per volume, hemoglobin per volume and hematocrit were decreased with respect to control samples (Figure 12, A). All these features describe the decrease of the global erythrocyte population due to the PK deficit. Alternatively, mean corpuscular volume (MCV) appeared increased on PKD samples, probably caused by the presence of immature erythrocytes. Based on the same rationale, immature erythrocytes, richer in hemoglobin content, increased the corpuscular hemoglobin (MCH) average of PKD samples compared with healthy donor and relative samples. All these traits correlate with chronic anemia compensated by reticulocytosis.[127] The same effect on MCV and MCH was observed in PKD mouse model (Figure 19). Finally, no differences were observed in the cell width distribution. Overall, these differences permitted the generation of ROC (Receiver Operating Characteristic) curves to set thresholds useful to distinguish between PKD patients and healthy donors (Figure 12, A: dotted line; B table).

A



B

	Area	P value	Thershold	Sensitivity %	Specificity %
Erythrocyte count	0.8867	<0.0001	< 4.34	80.95	90.63
Hemoglobin	0.8281	<0.0001	< 12.35	76.19	90.63
Hematocrit	0.8322	<0.0001	< 37.7	75	96.88
Mean corpuscular volume	0.801	0.0004	> 94.95	70	87.5
Mean corpuscular hemoglobin	0.8508	<0.0001	> 31.25	66.67	87.5

Figure 12. Erythroid parameters of PKD patients.
(**A**) Values concerning RBC subset are represented divided on healthy controls (C), patient´s relatives and PKD patients. Dotted line represents thresholds established based on ROC curves (n_C=32; n_{PKD}=20; $n_{Relatives}$=6; ****P<0.0001, ***P<0.001, *P<0.05) (**B**) ROC curves details based on selected thresholds. Thresholds were set trying to maximize specificity maintaining good sensitivity. Results represent mean values ± standard deviation.

Interestingly, relatives showed mean values between the ones found on healthy and deficient, but always remaining in the healthy side of the threshold set based on ROC curve. As expected, a not mutated allele was able to compensate the PKD phenotype. Nevertheless, this tendency if based on 6 samples, an insufficient number to consider it a robust data.

Myeloid subpopulations and lymphocytes were checked to find unbalances and its relation with PKD (Figure 13, A).

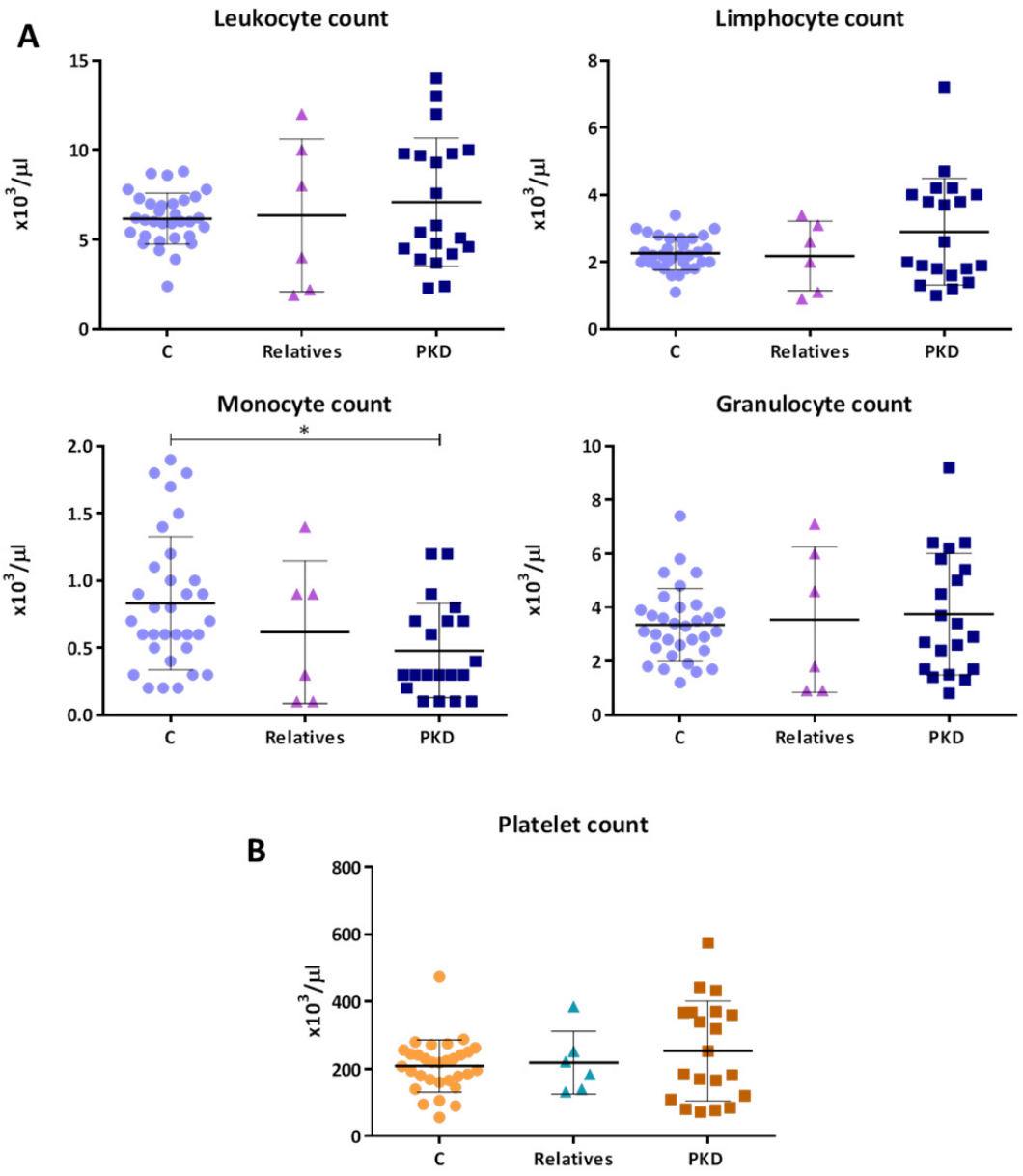


Figure 13. Leucocyte and platelet populations of PKD patients.
(**A**) Leucocyte populations measured by hematological counter (n_C=30; n_{PKD}=20; $n_{Relatives}$=6; *P<0.05) of healthy controls (C), patient´s relatives and PKD patients. (**B**) Platelet count comparing the three sample types. Results represent mean values ± standard deviation.

Among the three subsets measured, lymphocytes, monocytes and granulocytes, only monocyte counts showed altered levels with respect to the controls (0.8 ± 0.5 x 10^3/µL vs 0.5 ± 0.3 x 10^3/µL, respectively; mean±SD).

Finally, platelet population counts of PKD patient samples were no different from the controls and relatives (Figure 13, B).

To study deeper these alterations and to better characterize patient samples, a more detailed immunophenotyping study was carried out.

2. PKD patients immunophenotype

As reported, PKD affects only erythroid subsets. To further analyze the characteristics observed in the erythroid and myeloid subsets, a more detailed study was performed by flow cytometry.

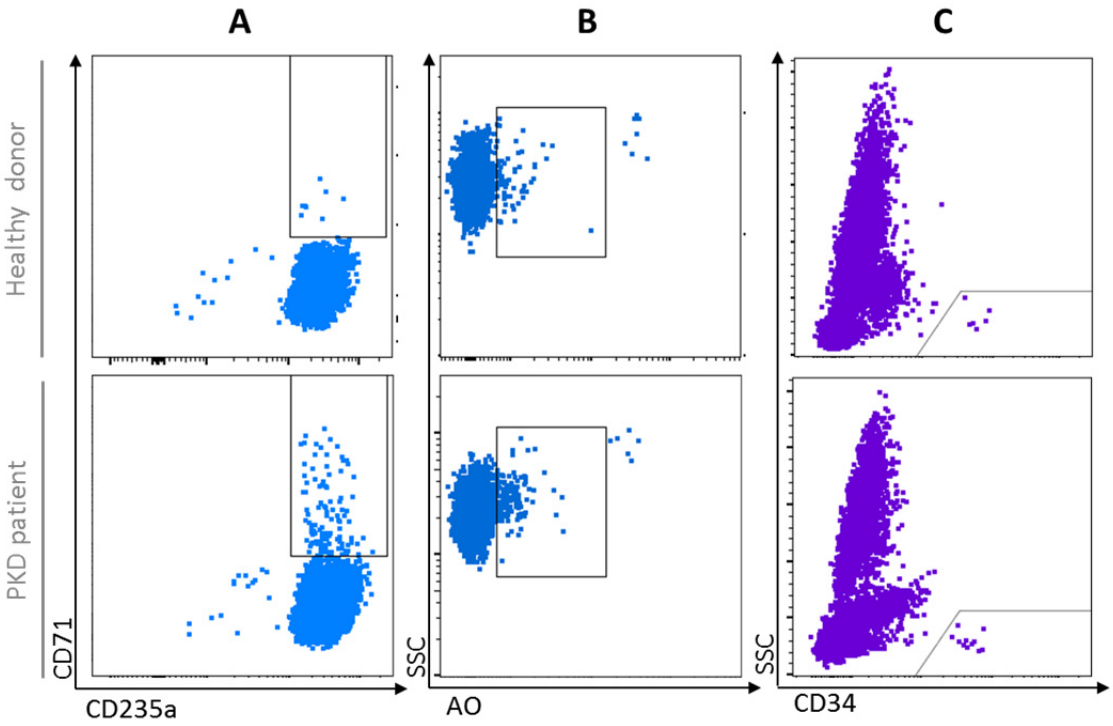
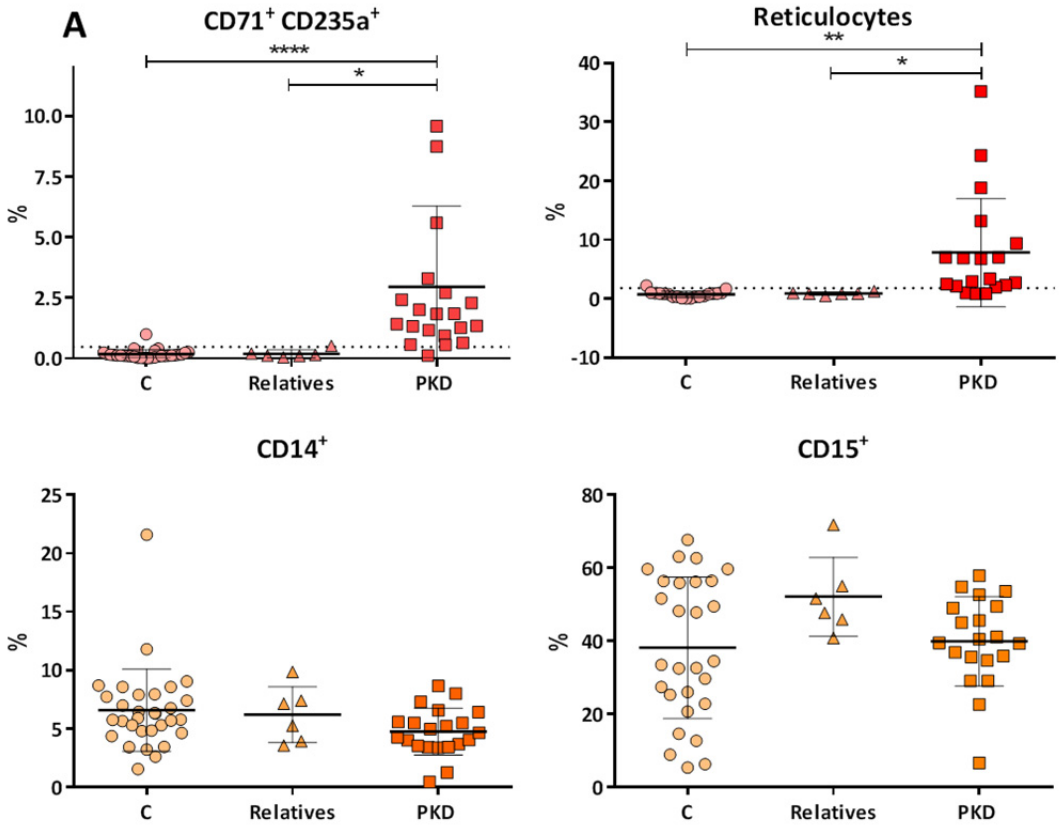


Figure 14. Representative dot plots of flow cytometry analysis of hematopoietic progenitor populations from healthy donors and PKD patients. (A) Erythroid progenitors. **(B)** Reticulocyte percentage based on acridin orange (AO) staining. **(C)** Hematopoietic progenitors.

First, immature erythroblasts, characterized by the expression of CD71 and CD235a antigens were analyzed (Figure 14, A). This cell subpopulation which is practically

inexistent in normal conditions (<0.17% in control cells) raised to 2.9% in studied PKD patients (Figure 15, A).

In the same way, reticulocyte percentages increased from 0.7% in controls to 7.8% in patient samples (Figure 14, B and Figure 15, A). Both increments are the response of the stressed bone marrow to compensate the anemia. Despite relatives´ hematological values (Hb, hematocrit, etc.) were between healthy and PKD patients, this was not the case for reticulocytes and progenitors where PKD relatives showed similar values to healthy donors. As performed in RBC counts, two ROC curves, one for each erythroblast population, were generated to set thresholds that discriminate healthy from PKD patients (Figure 15, A: dotted line; B table).



B

	Area	P value	Thershold	Sensitivity %	Specificity %
Reticulocytes	0.9662	<0.0001	> 1.795	84.21	94.74
CD71+ CD235a+	0.9391	<0.0001	> 0.475	95.24	96.77

Figure 15. Erythroid progenitors and mature myeloid populations measured by flow cytometry.
(**A**) Erythroid progenitors (CD71$^+$ CD235a$^+$), reticulocytes, monocytes (CD14$^+$) and granulocytes (CD15$^+$). Threshold set on erythroid progenitors (CD71$^+$ CD235a$^+$) and reticulocytes (dotted line) was based on

ROC curves described on (**B**) (n_C=30; n_{PKD}=20; $n_{Relatives}$=6; ****P<0.0001, **P<0.01, *P<0.05). Results represent mean values ± standard deviation.

Other myeloid subpopulations, monocytes ($CD14^+$) and granulocytes ($CD15^+$) were evaluated showing no statistical differences among PKD patients, relatives or healthy donors (Figure 15, A).

Moreover, classic lymphocyte populations were measured: B cells, followed by CD19, and T cells, tracked by CD3. Besides, helper T ($CD3^+$ $CD4^+$) and cytotoxic T ($CD3^+$ $CD8^+$) cells were also measured. No differences were found between healthy and deficient samples (Figure 16).

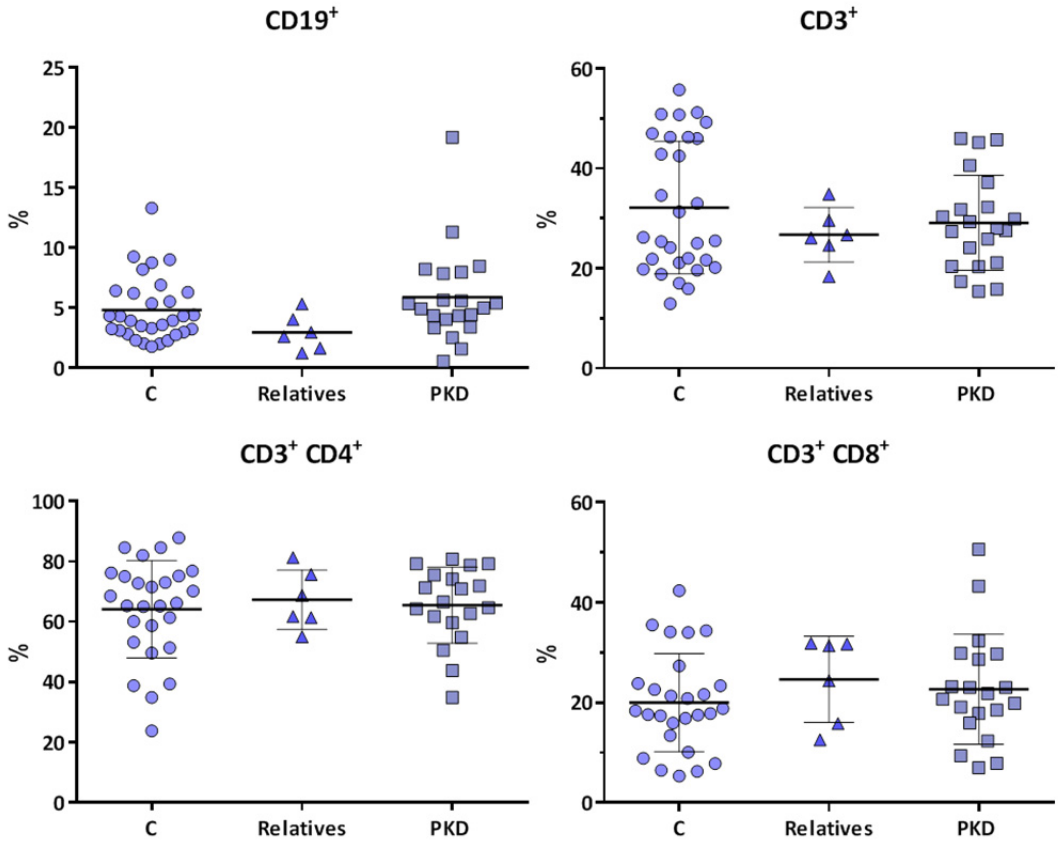


Figure 16. Lymphocyte populations measured by flow cytometry.
Cytometry percentages of healthy controls (C), patient´s relatives and PKD patients. B cells ($CD19^+$), T lymphocytes ($CD3^+$), T helper lymphocytes ($CD3^+$ $CD4^+$) and T cytotoxic lymphocytes ($CD3^+$ $CD8^+$; n_C=30; n_{PKD}=20; $n_{Relatives}$=6). Results represent mean values ± standard deviation.

3. Hematopoietic progenitors on peripheral blood

Flow cytometry was also used to evaluate the presence of hematopoietic progenitors circulating in peripheral blood (Figure 14, C). This subset is characterized by $CD34^+$ staining and it is normally barely represented on peripheral blood (lower than 0.1% of total cells). Although a slightly higher percentage of $CD34^+$ cells in peripheral blood was observed in PKD samples (control = 0.06 ± 0.04 %; PKD = 0.09 ± 0.07 %; mean±SD), there were no significant differences (Figure 17, A).

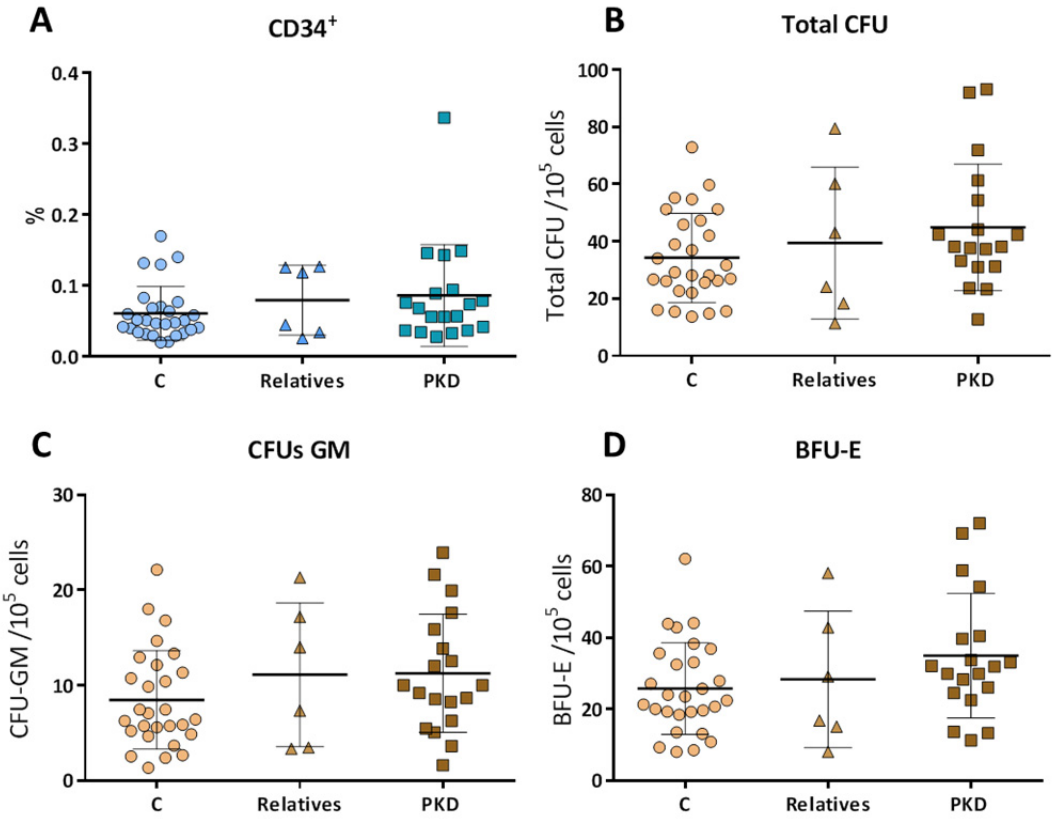


Figure 17. Hematopoietic progenitors present in peripheral blood.
(**A**) Hematopoietic progenitors present in peripheral blood of healthy controls (C), PKD patients and PKD relatives by flow cytometry. (**B**) Total colonies forming units generated from healthy controls (C), patient´s relatives and PKD patients. (**C**) CFU-GM: granulocyte-macrophages populations. (**D**) BFU-E: erythroid committed progenitors (n_C=28; n_{PKD}=19; $n_{Relatives}$=6). Results represent mean values ± standard deviation.

The same question was addressed analyzing the colony forming cell content. Mononuclear cells purified from peripheral blood were seeded on enriched semi solid medium doing triplicates, of 250,000 cells each, to allow the growth of hematopoietic

colonies coming from individual progenitors (CFUs). Also, erythropoietin was added to this enriched methyl cellulose medium to permit the visualization of erythroid progenitors. After 14 days, colonies were scored.

Although not statistically significant, PKD patient´s samples showed a slight increase in total colonies, about 1.3 times higher than the colonies formed by control samples (Figure 17, B). Interestingly, that tendency was also found in $CD34^+$ population measured by flow cytometry (Figure 17, A). This was also observed on the two different progenitor types identified, colony forming unit granulocyte-macrophages (CFU-GM; Figure 17, C) and erythroid burst forming unit (BFU-E; Figure 17, D).

Apart from the data showed before, other correlations were studied between severity of the disease (estimated by the hematological values) and other patient characteristics, such as mutations and gender.

Patients studied in this thesis showed more severe phenotype when they were double non-missense mutation (NM/NM) versus double missense mutation (M/M). For instance, hemoglobin values were: **NM/NM**= 7 g/dL (n=1, mean±SD), **NM/M**=10±2 g/dL (n=4, mean±SD) and **M/M**=12±3 (n=11, mean±SD). Nonetheless, the number of patients per group was too low, so a higher number of samples would be required to confirm this correlation.

Finally, no significant differences were found comparing severity, represented by hemoglobin level, of different genders: **male** PKD patients = 11±3 g/d (n=11, mean±SD), **female** PKD patients = 12±3 g/d (n=9, mean±SD).

Definition of the minimal dose of corrected cells to cure PKD

To facilitate the applicability of gene therapy for PKD, the definition of the minimum corrected cell number required to cure the disease is important to know the optimal amount of transduced cells needed to be transplanted in PKD patients. Pre-clinical studies demonstrated the curative capacity of coRPK lentiviral vector for PKD in a PK deficient mouse model (AcB55). In this previous work, all cured mice showed VCN values higher than 1 vector copy per cell.[124] Nevertheless, the minimum amount of therapeutic vector required to compensate the disease was a question that remained unclear.

The following experiment series had the aim of elucidating the VCN threshold above which PKD mice show a healthy phenotype.

1. *In vivo* PKD correction and its relation with VCNs

Similarly to the eventual human clinical trial, the basic scheme of the experiments was based on *ex vivo* correction of hematopoietic progenitors and their transplantation into conditioned PKD mice.

Thus, mouse hematopoietic progenitor cells (Lin^-) were sorted from total bone marrow samples extracted from male PKD mice and transduced as explained (see *Hematopoietic progenitor transduction protocol, Materials and Methods*). Two lentiviral vectors were used, the therapeutic coRPK vector and the EGFP vector as control. After a 24 hours pre-stimulation period, two transduction rounds (Multiplicity of infection, MOI=10) were performed after 0 and 24 hours post pre-stimulation. Six conditioned female mice (3 per treatment; two doses of 4.5 grays) were transplanted with 400,000-500,000 progenitor cells/ mouse.

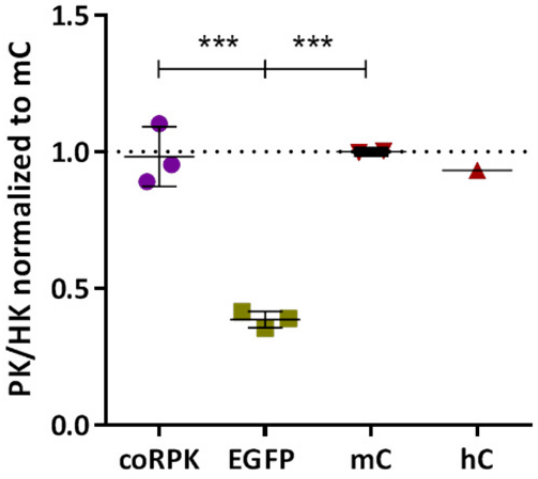
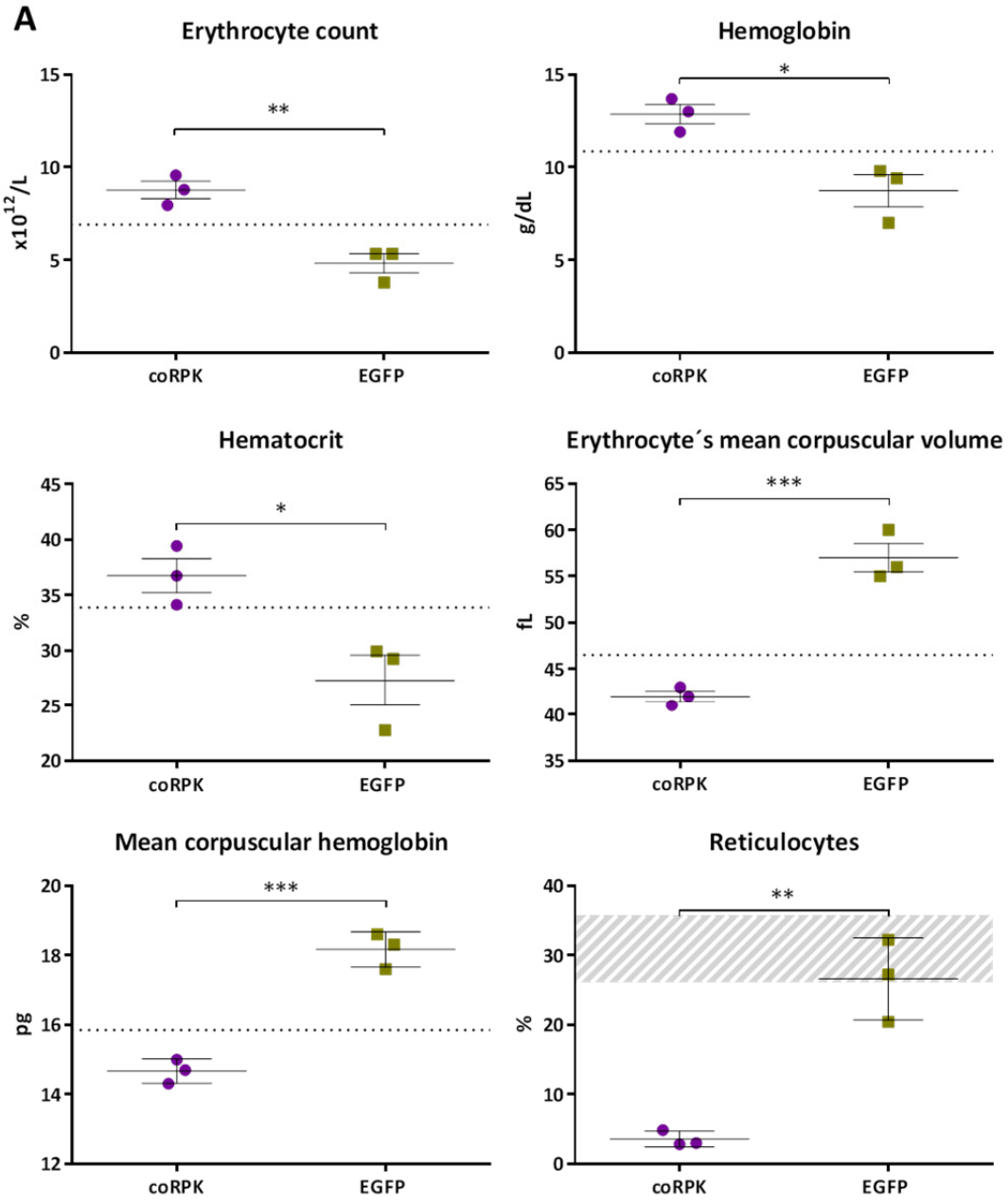


Figure 18. PK/HK activity from RBC of mice transplanted with transduced hematopoietic progenitors. mC: healthy control mouse cells (n=2). hC: healthy control human cells. Measures were normalized with respect to the mean of mC samples. (coRPK and EGFP n=3 each; ***P<0.001). Results represent mean values ± standard deviation.

Forty days post-transplant, peripheral blood samples were taken to assay pyruvate kinase activity (Figure 18). Mice transduced with therapeutic vector recovered enzymatic activity levels comparable to those measured in control mice and human healthy donors, while PKD mice transduced with the EGFP vector remain under 50% of activity.



B

	Area	P value	Thershold	Sensitivity %	Specificity %
Erythrocyte count	1	<0.0001	< 6.9	93.75	92.31
Hemoglobin	0.95	<0.0001	< 10.85	87.5	92.31
Hematocrit	0.9	<0.0001	< 33.88	92.68	83.78
Mean corpuscular volume	1	<0.0001	> 46.5	97.5	100
Mean corpuscular hemoglobin	0.93	<0.0001	> 15.85	92.68	91.89

Figure 19. Hematological features of PKD mice transplanted with corrected hematopoietic progenitors one year post-transplant of genetically corrected cells.

(**A**) Hematological parameters of mice transduced with coRPK (purple circles) and control EGFP vector (green squares). Dashed interval in reticulocytes graph represents standard deviation of control deficient mice (AcB55). Dotted horizontal lines are the thresholds set based on ROC curves, which values are represented on table (**B**; ***P<0.001, **P<0.01, *P<0.05). Results represent mean values ± standard deviation.

The experiment was followed up for a year. At the end point, spleen, bone marrow and peripheral blood were collected. The study of the hematological parameters of peripheral blood at end point revealed a complete correction of the deficient profile (Figure 19, A). Erythrocyte count, hemoglobin and hematocrit of corrected mice recovered values above the threshold stablished by the ROC curves (generated from healthy, C57BL/6, and PKD, AcB55, mice samples; Figure 19, B).

Also, the volume of the erythrocytes and the amount of hemoglobin per cell decreased to normal levels indicating a rescue from the anemic situation (Figure 19, A). Correlating with these results, the corrected mice did not present reticulocytosis (Figure 19, A).

Spleen size is a good indicator of the rescue of the disease. PKD mice suffer splenomegaly due to the increased splenic activity that have to process all deficient erythrocytes. Mice treated with coRPK vector reduced its spleen size and weight to normal values (Figure 20).

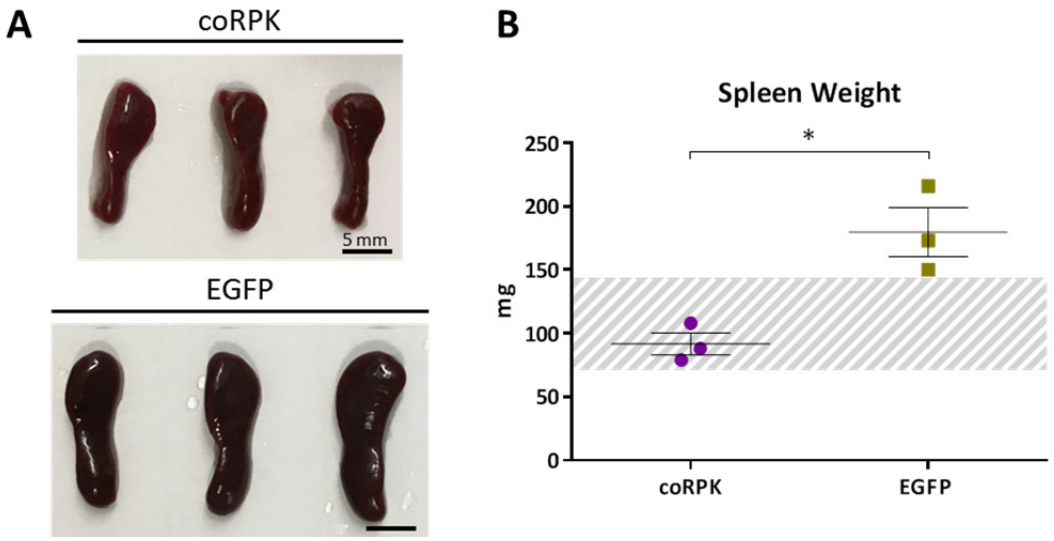


Figure 20. Spleen size and weight of mice transplanted with control or corrected hematopoietic progenitors.
Size **(A)** and weight **(B)** of mice´s spleens transduced with therapeutic vector (coRPK) and control vector (EGFP). Dashed interval represents the healthy zone based on the standard deviation of control healthy C57BL/6 mice (*P<0.05). Results represent mean values ± standard deviation.

Bone marrow samples were taken to elucidate the transduction efficacy and the number of transgene copies that produced this correction. Cells extracted from bone marrow were cultured on methyl cellulose for 7 days.

DNA from CFUs was analyzed by qPCR to measure the VCN in each individual colony. The analysis of the transduced/non-transduced CFUs showed that all mice within coRPK group had near 100% transduced CFUs. In the case of EGFP mice, one animal had near 100% transduction while the others had around 70% (Figure 21, A).

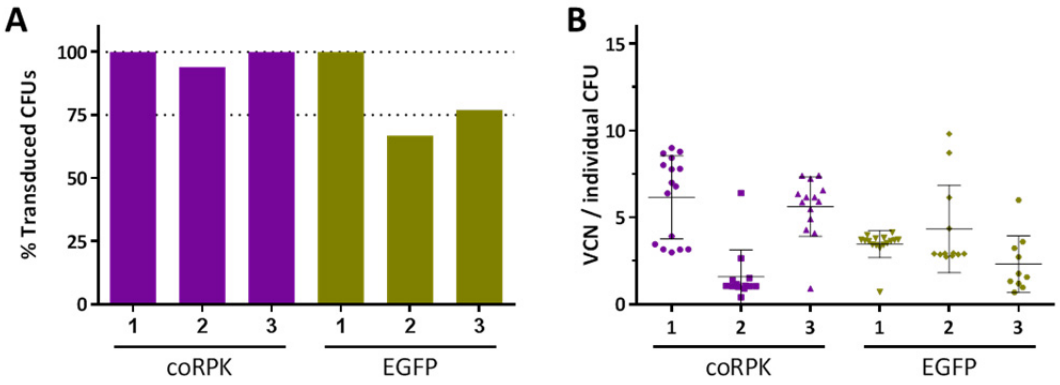


Figure 21. Transduction efficacy and VCN in colonies forming units from bone marrow of mice transplanted with control or corrected hematopoietic progenitors.
(**A**) Percentage of CFUs positive for the transgene among total colonies evaluated by qPCR. (**B**) VCN evaluated by qPCR of transduced colonies. Number of colonies measured=13-16. Colonies were considered positive when VCN was higher than 0.15 copies /cell. Results represent mean values ± standard deviation.

The technique used is based on counting the proportion of transduced/non-transduced CFUs. To confirm the reliability of the qPCR method, the results obtained were compared with the percentage of EGFP positive hematopoietic cells in bone marrow and in spleen by flow cytometry. Percentage of transduced cells taking into account all the three methods together for each EGFP mouse was 83±17%, 49±17% and 75±4%, respectively. CFU transduction proportion measured by qPCR and flow cytometry of the bone marrow and spleen data were consistent.

VCN was evaluated individually on each colony by qPCR (Figure 21, B). Despite of the fact that the cells transplanted on each mouse came from the same pool of transduced cells, VCN slightly varied among the different mice. Nevertheless, all mice transduced with

coRPK were corrected meaning that at least 2±1 copies per cell were enough to correct PKD phenotype (Figure 21, B).

With the aim of precisely defining the minimum number of corrected cells to have a therapeutic effect and the correcting VCN limit where PKD mice are not capable of rescuing the disease, additional experiments were performed covering a large MOI interval. The idea was to be able to correlate VCN with either corrected or deficient mice. Thus, MOIs from 0.3 to 50 were tested in a new experimental series.

2. Minimum therapeutic VCN

The general scheme of this experiment was similar to the previous one, but here, to avoid variations, a unique transduction cycle was performed. Moreover, to make the results independent of transduction efficacy and engraftment percentage, the final read out was done taking into account the VCN. The experimental series were done in two independent experiments, one at high MOIs (25 and 50) and another covering a wide spectrum of low MOIs (0.3, 1, 3, 10, 25), having MOI=25 group in common.

For high MOIs experiment, seven male PKD (AcB55) mice were sacrificed and Lin^- cells were sorted from their bone marrow. After one transduction round with coRPK (MOIs: 25, 50) or EGFP (MOI=100), cells were transplanted in conditioned female PKD mice ($1x10^6$ cells/mouse, 3 mice per each group; two doses of 4.5 grays). VCN and hematological parameters were followed for 215 days (Figure 22).

Reticulocyte percentages of control EGFP mice were mostly two times higher than the values obtained in the coRPK transduced ones all over the experiment, indicating that mice transduced with the therapeutic vector were corrected (Figure 22). The egress of reticulocytes showed a characteristic dynamic: a sharp increase around day 70 followed by stabilization, close to healthy values in the case of coRPK mice (Figure 22).

With respect to the other erythroid parameters, coRPK mouse scores remained around the healthy values all over the experiment, while EGFP values fitted with PKD values (Figure 22).

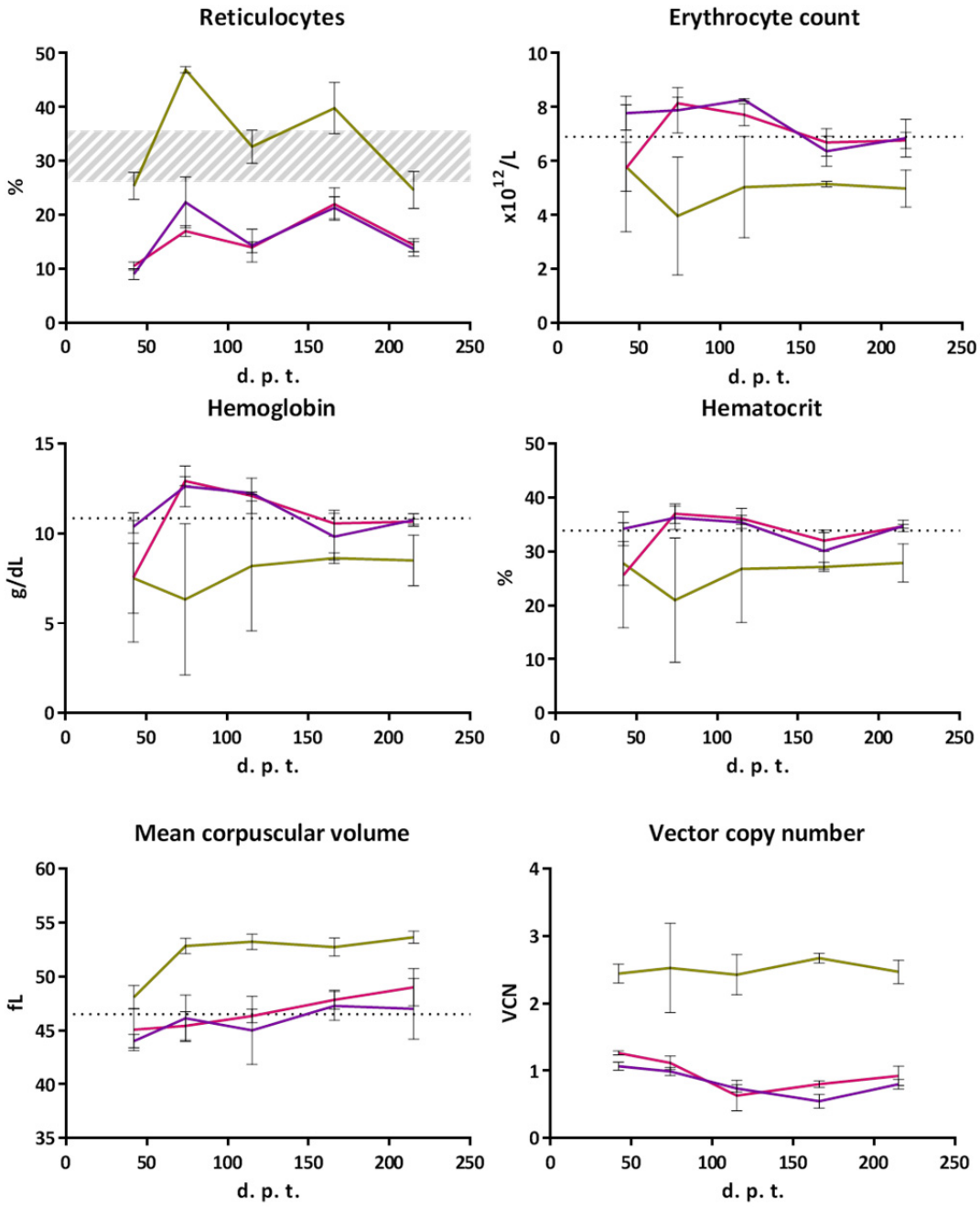


Figure 22. Erythroid parameters and VCN over time of mice transplanted with control or corrected cells and analyzed up to 215 days post-transplant (d. p. t).
Green line: mice transduced with EGFP vector. Purple line: mice transduced with coRPK MOI 50. Pink line: mice transduced with coRPK MOI 25. Dashed interval in reticulocytes graph represents standard deviation of control deficient mice (AcB55). Dotted horizontal lines are the thresholds set based on ROC curves, which values are represented on table B, Figure 19. Results represent mean values ± standard deviation.

VCN was evaluated by qPCR from peripheral blood samples along the experiment. MOI 25 and 50 mice presented similar and very stable copy/cell values despite the different number of viral particles used. For instance, at day 42 post-transplant, MOI 25 VCN was

1.07 ± 0.06 meanwhile MOI 50 was 1.27 ± 0.03. Despite VCN fluctuated over time, they always continued around 1 copy per cell (Figure 22).

In the next experiment, lower MOIs were tested (0.3-25; Figure 23).

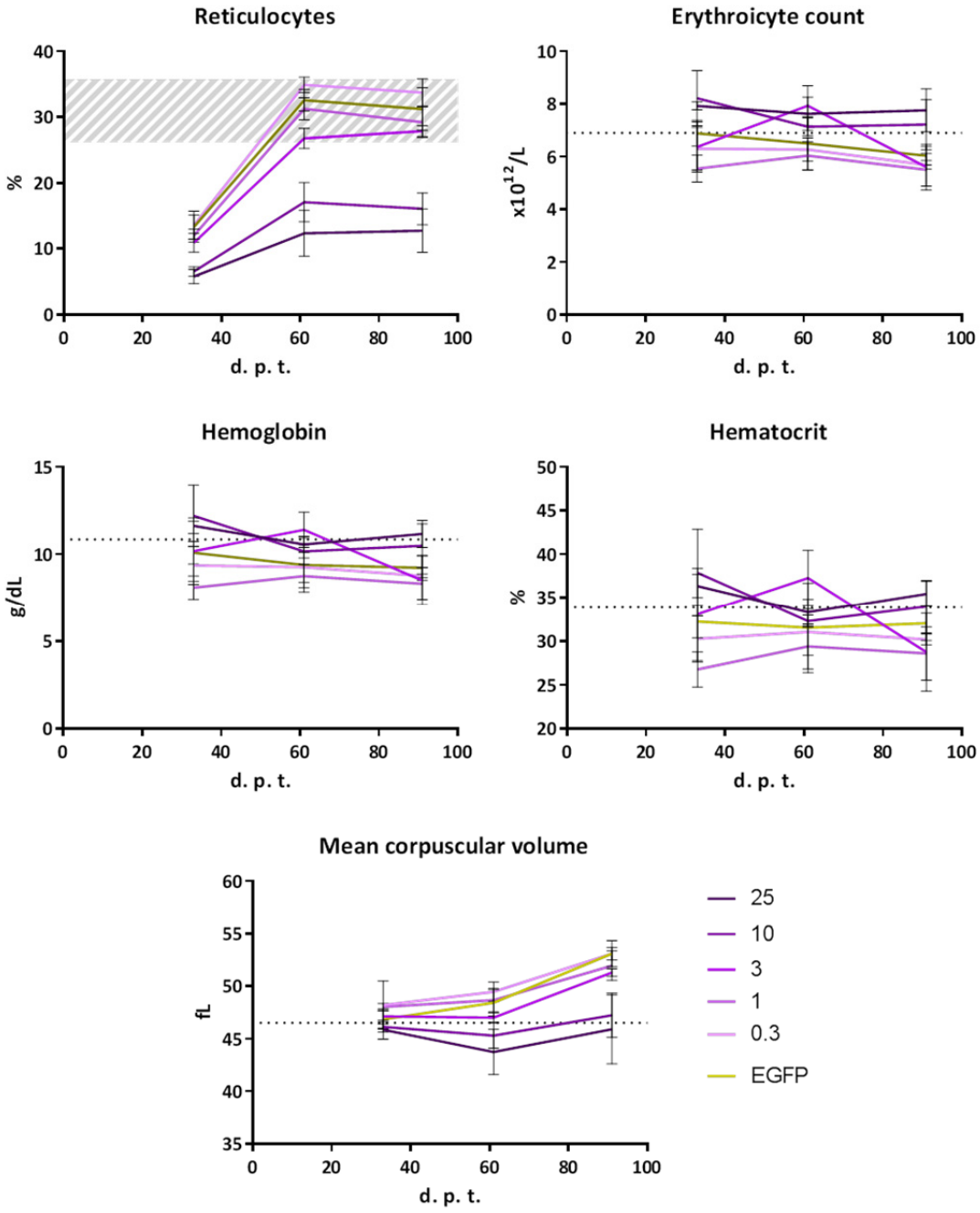


Figure 23. Hematological parameters over time of mice transplanted with cells transduced with MOIs from 0.3 to 25 of coRPK or EGFP.

Results are represented up to 90 days post-transplant (d. p. t.). Green line: mice transplanted with cells transduced with EGFP vector. Purples lines: mice transplanted with cells transduced with coRPK vector, the darkest line represents the highest MOI, 25, while the lightest line represents the lowest, 0.3. Dashed interval in reticulocytes graph represents standard deviation of control deficient mice (AcB55).

Dotted horizontal lines are the thresholds set based on ROC curves, which values are represented on table B, Figure 19. Results represent mean values ± standard deviation.

Lin$^-$ PK deficient progenitors were sorted and transduced with coRPK (MOIs: 0.3, 1, 3, 10 and 25) or EGFP (MOI=25). After transduction, cells were transplanted in conditioned female PKD mice (1x10^6 cells/ mouse, 3 mice per each group, except groups MOI 1 and 3 groups in which 4 mice per group were transplanted). Hematology parameters were followed for 90 days showing correction differences since day 30.

MOI 10 and 25 groups showed correction at day 60 while the rest groups remained at higher reticulocyte values after a decrease observed at 30 days in all groups, probably due to the effect of the hematopoietic conditioning and transplant (Figure 23, reticulocytes. Dashed area represents deficient values). Efficacy of the treatment depends not only on the MOI used but also of the percentage of engraftment of the transplanted cells reached. Thus, correlation of phenotype correction with VCN was only possible being independent of these two parameters.

VCNs of both experiments were evaluated and compared to see the correlation with respect of the MOI used (Figure 24).

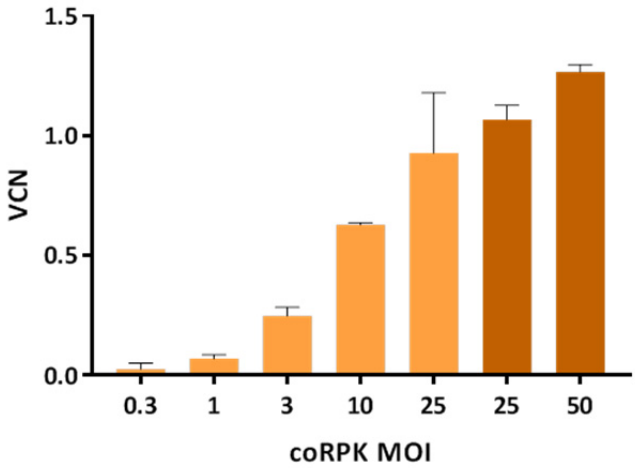


Figure 24. Correlation between VCN reached in the animals with respect to coRPK MOI used in both experiments.
Light color columns represent results of the low MOIs experiment (day 30 post-transplant). Dark color columns represent results of the high MOIs experiment (day 42 post-transplant). Results represent mean values ± standard deviation.

As expected, VCN obtained increased progressively with MOI. Moreover, the MOI 25 group, in common in both experiments, showed comparable VCN values (Figure 24).

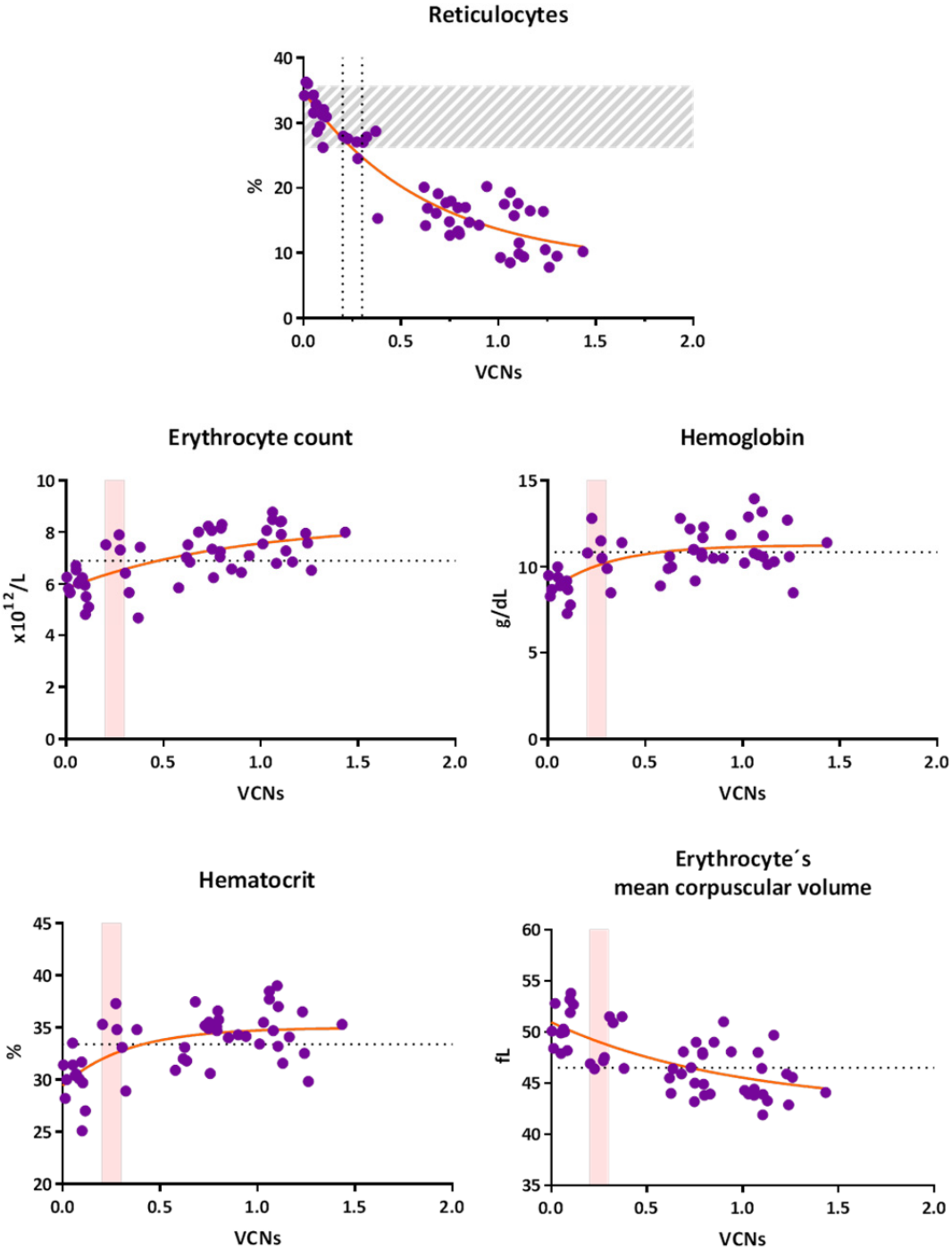


Figure 25. VCN versus hematological parameters.
VCNs obtained in both experiments together. Up: Dotted vertical lines correspond to 0.2 and 0.3 VCNs, interval which fits with the minimum therapeutic VCN. Down: Pink columns correspond to 0.2 and 0.3 VCNs. Orange line, non-lineal correlation. Dashed interval in reticulocytes graph represents standard deviation of control deficient mice (AcB55). Dotted horizontal lines are the thresholds set based on ROC curves, which values are represented on table B, Figure 19.

Finally, to address the question of minimum number of copies per cell to correct the disease, VCNs obtained along all sampling points of both experiment were plotted against the different hematological parameters (Figure 25).

As happened in the graphs that show the hematological parameters over time, here it is also reticulocyte percentage where a clear correlation could be observed (R^2=0.868), being VCNs between 0.2-0.3 the limit below which mice could not be rescued. The rest of parameters also showed PKD phenotype below 0.3 VCN but the disperse distribution of the values did not permitted a good correlation (R^2<0.5).

Thus, thanks to these correlations a minimum copy number of coRPK per cell in peripheral blood was stablished to cure PKD, being between 0.2 and 0.3.

PKD model

As it has been shown in *Patients characterization* section, immunophenotyping of patient´s cells is possible using small amounts of peripheral blood. However, peripheral blood is a poor source of PKD hematopoietic progenitors that are needed to test and improve new advanced cell and gene therapies for this disease. From a standard sample of 3 ml of peripheral blood barely a few thousands $CD34^+$ cells can be obtained. For a deeper study of the disease and to have material enough to test different curative approaches, a new source of PKD progenitors is needed. To beat this drawback, PKD-like cells could be generated by modifying the gene expression or the genome of healthy hematopoietic progenitor cells.

Hence, to address this issue two different approaches were applied. The first was based on the use of shRNAs recognizing specific sequences of the RPK mRNA packaged into lentiviral vectors. The second approach was based on the transitory electroporation of the CRISPR/Cas9 system with specific guides for *PKLR* gene. Both strategies are meant to decrease RPK levels, disrupting either mRNA or directly the progenitor cell genome, respectively.

1. PKD-like cells generated by shRNAs

1.1 *In vitro* RPK silencing in K562 cell line

The aim of the first strategy was to generate PKD-like cells by silencing RPK mRNA expression. As explained in *Materials and Methods*, six shRNAs were designed to recognize

the wt RPK mRNA sequence and not the therapeutic vector sequence (coRPK). A first screening was done with three of the six shRNAs, Sh1, Sh3 and Sh5, targeting exons 11, 1 and 6, respectively. The designed shRNAs were cloned in a lentiviral vector backbone that also expressed the EGFP protein to easily trace transduced cells. As control, a scramble shRNA viral vector not recognizing any specific cellular mRNA (Scrb) was developed and used in parallel. To validate shRNAs efficacy, a cell line from a human chronic myelogenous leukemia (K562) that expresses the RPK isoform was used.

About 80,000 cells for each condition were transduced (MOI=10) with the three different viral vectors expressing shRNAs and Scrb. After amplifying the culture, cells were collected to evaluate RPK level by western blot.

While all shRNAs were able to decrease RPK level, Sh1 and Sh3 seemed to be the most effective in reducing RPK production. After Sh1 transduction, RPK was not visible on western blot membrane and Sh3 transduction reduced about 80% RPK level compared with control Scrb sample (Figure 26). Once these good results were obtained, these three shRNAs were selected to continue the screening, discarding Sh2, Sh4 and Sh6.

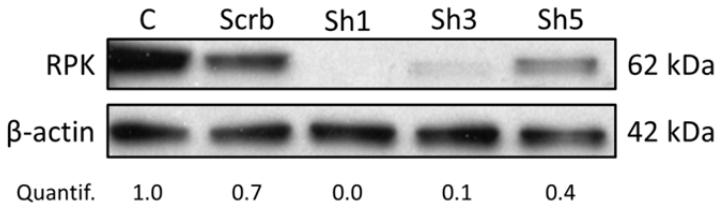


Figure 26. Western blot of K562 cell line interfered with different RPK shRNAs.
K562 cell line expresses endogenous RPK protein. C: non modified K562, non-transduced cells; Scrb: cells transduced with scramble vector; Sh1, Sh3 and Sh5, the three different shRNA vectors tested (80,000 cells/condition; MOI=10). β-actin was used as loading control. Quantification was performed normalizing with respect to β-actin and with respect to non-modified K562 (C).

Another basic requirement to select one of these shRNAs was that they should not recognize the codon optimized sequence of coRPK.

To corroborate this, different cell lines not expressing RPK were first transduced with the therapeutic vector (coRPK) and then with the different shRNA-LVs. HL-60 cells, a human promyelocytic leukemia cell line, was transduced with coRPK (MOI=10) and,

afterwards, transduced with shRNA 1, 3 and 5 (MOI=10). Transduced cells were sorted based on the expression of the EGFP reporter expression.

None of the shRNAs tested were able to prevent coRPK production, meaning that any shRNA affected coRPK expression (Figure 27).

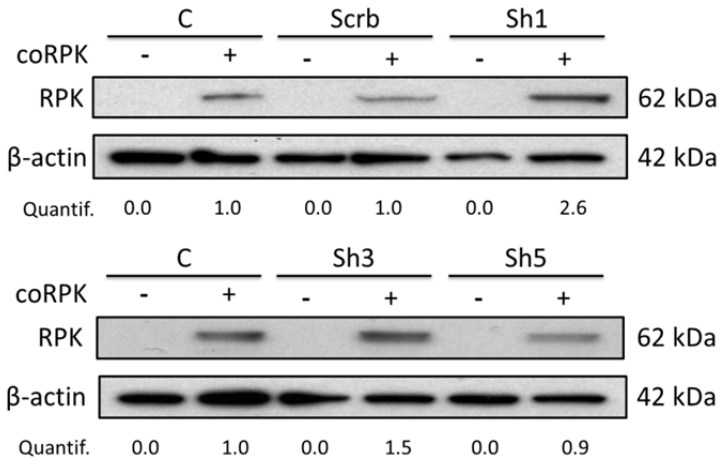


Figure 27. Western blot of HL-60 cell line interfered with different RPK shRNAs after being transduced with coRPK transgene.
+: cells transduced with coRPK vector (MOI=10); C: control, non-transduced cells; Scrb: cells transduced with scramble vector; Sh1, Sh3and Sh5: the three different shRNAs tested, (MOI=10). β-actin was used as loading control. Quantification was performed normalizing with respect to β-actin and with respect to non-modified HL-60 (C).

To assure that shRNAs were not capable of silencing coRPK, a second cell line was transduced at different MOIs of coRPK. HT1080 cells (a routine cell line which does not express endogenous RPK) was transduced with decreasing MOIs of 60, 6 and 0.6 of coRPK and again transduced with the most potent shRNA (sh1) at constant MOI of 10. RPK expression from coRPK vector was visible even in the lowest MOI tested, meaning that Sh1 did not affect transgene protein production (Figure 28).

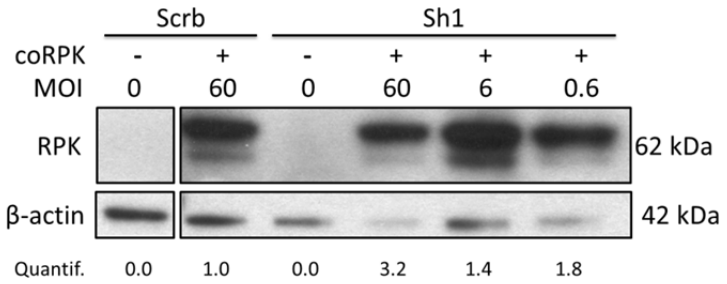


Figure 28. Western blot of HT1080 cell line transduced with different MOIs of coRPK transgene and interfered with Sh1.
Scrb: cells transduced with scramble vector; Sh1: MOI=10. coRPK: MOIs of 60, 6 and 0.6. Quantification was performed normalizing with respect to β-actin and with respect to non-modified HT1080 (C).

1.2 *PKLR* gene silencing in hematopoietic progenitors

The next step was to study if the developed shRNA system to silence RPK expression was efficient in primary human hematopoietic progenitors. *In vitro* cell line testing showed Sh1 as the most potent shRNAs. Therefore, it was the one chosen for these studies. Healthy hematopoietic progenitors ($CD34^+$) obtained from cord blood were transduced with the Sh1 or the Scrb lentiviral vector and put into an *in vitro* erythroid differentiation protocol.

A 2-3 weeks erythroid differentiation protocol was used in which different cytokine cocktails promoted the hematopoietic progenitors differentiation to become mature erythroblasts. In these cultures, RBC maturation can be followed by changes in surface markers. CFU-E are $CD36^+/CD45^+/CD71^+/CD235a^-$. Progressively along maturation, erythroblasts lose CD45, CD36, later they gain CD235a and finally they lose CD71 (Table 9, A).[128,129]

However, the *in vitro* differentiation protocol used here changed slightly the timing of the decrease of some surface markers with respect to *in vivo* normal differentiation. Normally, CD45 and CD36 are the first two antigens that pro-erythroblasts lose.[129] Thus, when $CD71^+$ reach its highest expression, cells are completely negative for CD45 and mostly for CD36 (Table 9, A, I). After reaching maximum CD235a expression, differentiation stages are divided in three subpopulations based on CD71 decrease (Table 9, A, II-IV).

Table 9. Erythroid maturation immunophenotypes.

A. Immunophenotype along ***in vivo*** erythroid maturation

	CD235	CD71	CD36	CD45
I	-	+	+	+
II	+	+	mid	-
III	+	mid	low	-
IV	+	-	-	-

B. Immunophenotype along ***in vitro*** erythroid maturation

	CD235	CD71	CD36	CD45
I	-	+	+	+
II	+	+	+	mid
III	+	mid	mid	low
IV	+	-	-	-

In the *in vitro* differentiation used, CD36 and CD45 expression could be found within the $CD71^{+}$ population. To illustrate this, the immunophenotype of three different samples is shown Figure 29.

The first two columns are primary cells that were obtained from a mobilized peripheral blood apheresis of a healthy donor and from cord blood. In both cases, most cells were already CD36 and CD45 negative. Populations II (orange), III (purple) and IV (red) were established depending on CD71 content (Table 9, A). On the other hand, the third sample corresponds with an *in vitro* erythroid differentiation of primary $CD34^{+}$ cells. The CD36/CD45 dotplots showed a clear transition from $CD45^{mid}/CD36^{+}$ to $CD45^{-}/CD36^{-}$ (Table 9, B, red square). Additionally, this transition fitted with the decrease of CD71 (Figure 29, right side). Another difference was that most cells found on the first two samples, followed by the side histograms, were already $CD71^{-}$ (IV) while in the *in vitro* differentiation the main cell population was subpopulation III (Figure 29, histograms).

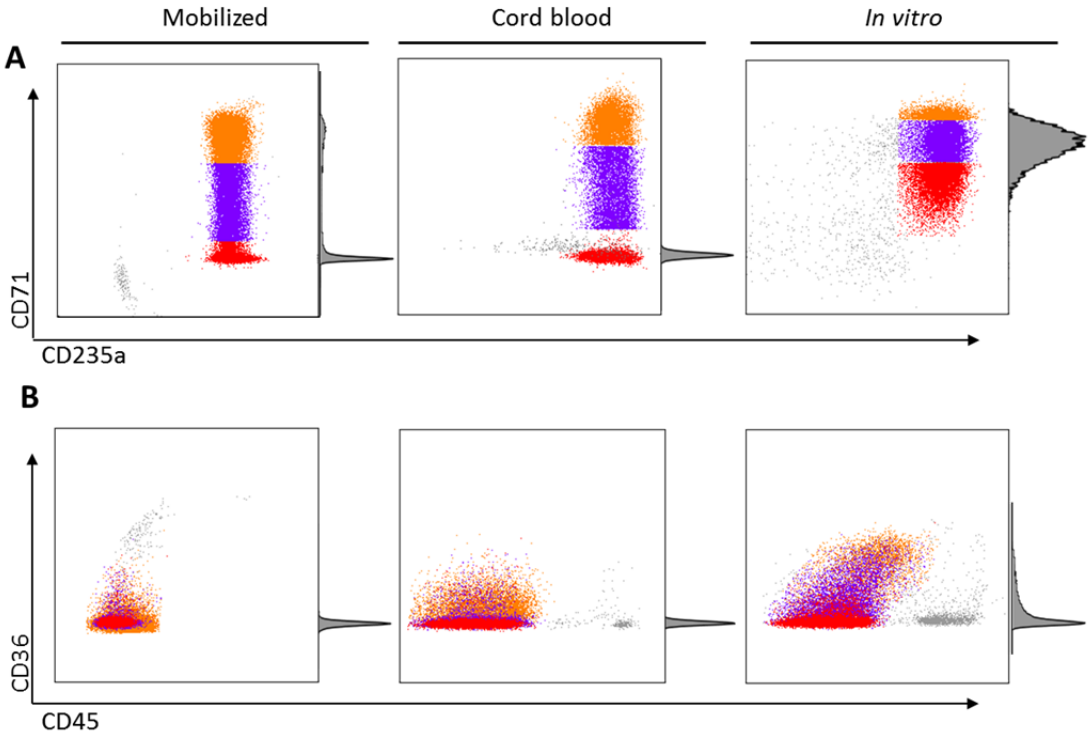


Figure 29. Erythroid lineage profile of *in vivo* and *in vitro* samples.
Mobilized: sample obtained from apheresis of healthy mobilized peripheral blood. Cord blood: erythroid lineage obtained from healthy cord blood. *In vitro*: erythroid lineage obtained from *in vitro* differentiated after 14 days from hematopoietic progenitors obtained from cord blood (**A**) CD71 decrease over maturation with respect to CD235a population. (**B**) CD36/CD45 decrease over maturation. Right-side histogram: cell distribution among total cells on each sample. Erythroid subpopulations are represented in orange (II), purple (III) and red (IV).

Therefore, the criterion followed for *in vitro* erythroid maturation was not based just on CD71 decrease but also on CD36 and CD45 changes. (Table 9, B). Thus, II, III and IV *in vitro* subpopulations do not correspond exactly with their *in vivo* homologues, but in both cases represent the same erythroid maturation stages. Hereinafter, II, III and IV sub-populations must be understood as *in vitro* subpopulations.

After 14 days, cell cultures obtained from the $CD34^{+}$ cells transduced with the interference lentiviruses were already composed by erythroid cells and its differentiation stage was evaluated by immunophenotype.

Flow cytometry analysis was based on the following gating strategy. First, GFP^{+} (interfered) and GFP^{-} (non-interfered) populations were gated separately (Figure 30, A). Then, $CD36^{mid/-}$ $CD45^{mid/-}$ populations were selected (Figure 30, B). Afterwards, following the criterion previously explained (Table 9, B), populations II

(CD235a^{+}/CD71^{+}/CD36^{+}/CD45^{mid}), III (CD235a^{+}/CD71^{mid}/CD36^{mid}/CD45^{low}) and IV (CD235a^{+}/CD71^{-}/CD36^{-}/CD45^{-}) were gated on a CD71/CD235a dot-plot (Figure 30, C). This was finally back-gated on the CD36/CD45 plot following the same color code (Figure 30, B).

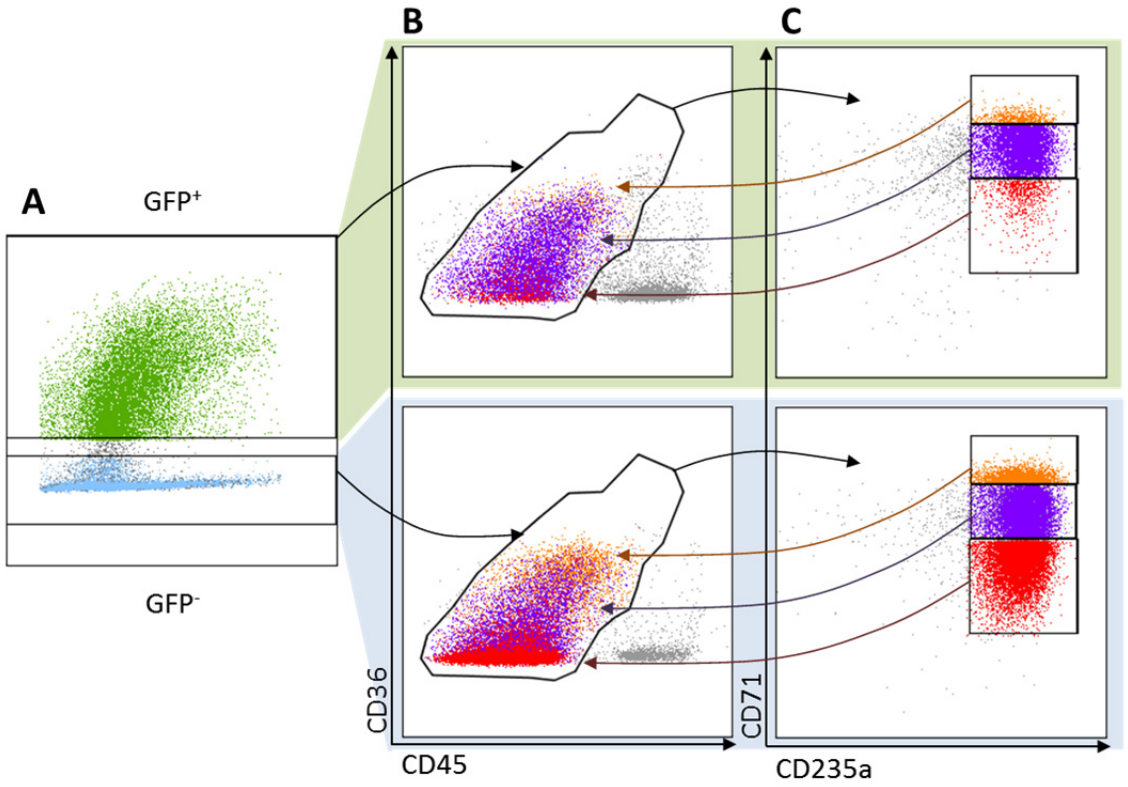


Figure 30. Gating strategy of transduced hematopoietic progenitors after differentiated to erythroid lineage.
Progenitors differentiated after Scrb transduction have been used as example on this scheme (**A**) GFP positive and negative selection: erythroid maturations of both populations were analyzed separately. (**B**) CD36/CD45 dot-plots. (**C**) CD71 vs. CD235a where II (orange), III (purple) and IV (red) subpopulations were gated and then back-gated in (**B**).

Both healthy (GFP^{-}) and deficient (GFP^{+}) cells could be differentiated along the erythroid lineage and studied under exactly the same conditions.

Erythroid profile of cells transduced with Sh1 and Scrb lentiviruses were followed, being Sh1^{+} and Scrb^{+} the GFP positive fraction of each culture and Sh1^{-} and Scrb^{-} their GFP negative counterparts (Figure 31, A and B).

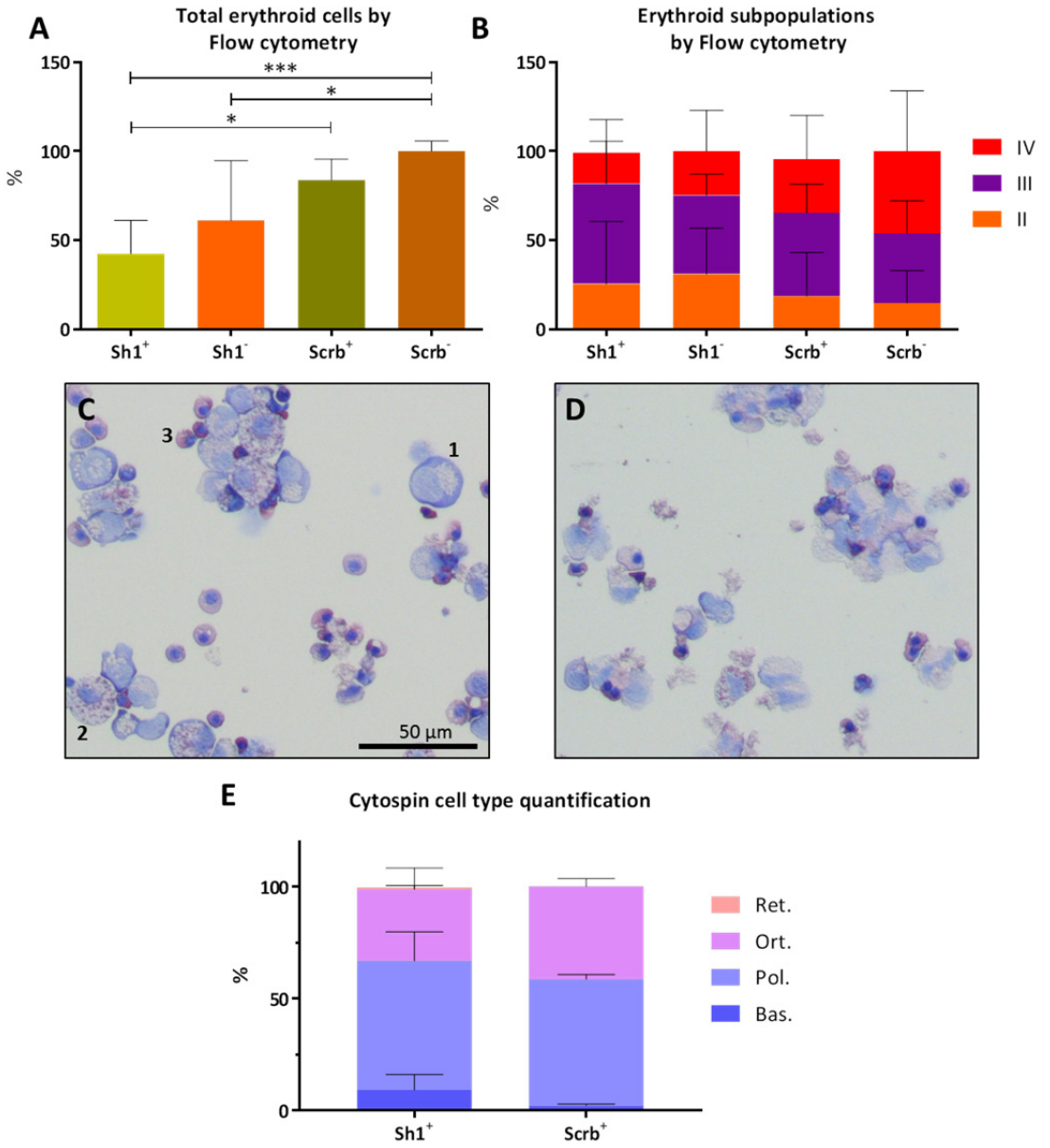


Figure 31. Hematopoietic progenitors transduced with either Sh1 or scramble (Scrb) lentiviral vector and differentiated to erythroid lineage.
(**A**) Total erythroid cells (CD235a$^+$ within CD36/CD45, Figure 30, B) out of total cells. Sh1$^+$ and Sh1$^-$ are the GFP$^+$ and GFP$^-$ fractions of cell culture transduced with Sh1. Scrb$^+$ and Scrb$^-$ correspond with GFP$^+$ and GFP$^-$ fractions of cells transduced with the scramble shRNA control. These populations include II, III and IV fractions. Measures are based on cytometry percentages and normalized with respect to Scrb$^-$ percentage, which is consider =100% (***P<0.001, *P<0.05; n=9). (**B**) Erythroid maturation stages II (orange), III (purple) and IV (red) percentages within CD235a$^+$ population (n=9). (**C**-**D**) Representative cytospin micrographs from cultures after Sh1 (**C**) or Scrb (**D**) transduction and differentiation. Pictures show GFP$^+$ fractions that were sorted to isolate transduced cells. (**E**) Quantification of cell types present on cytospin preparations distinguishing basophilic erythroblast (C, 1) polychromatophilic erythroblast (C, 2), orthochromatic erythroblast (C, 3) and reticulocytes (Sh1 n=2, Scrb n=2).

Small variations among the different experiments were observed due to differences in the efficacy of erythroid differentiation. To correct these technical variations, each

sample was normalized with respect to the Scrb$^-$ samples (in theory, the most similar to the wild type) of each experiment.

Total Sh1$^+$ cells that reached erythroid lineage were less than 50% obtained transducing with Scrb vector (Figure 31, A). This decrease can fitted with the metabolic alteration induced because of *PKLR* gene silencing. On the other hand, a higher erythroid population was expected on Sh1$^-$ population, as long as they were not interfered (Figure 31, A).

Regarding the hypothesis that PKD could unbalance erythroid II-IV subpopulations, the different transduced cells were compared. If PK activity is crucial for moving towards the final erythroid stages, an accumulation of subpopulations II or III would be expected. After flow cytometry characterization, significant differences could not be found among II-IV subpopulations. Despite Scrb$^-$ cells showed a tendency of having higher IV proportion than to Sh1$^+$ population, there were not statistical differences among any of the samples, either GFP$^+$ or GFP$^-$ (Figure 31, B). This lack of differences was also observed in cytospin samples. Before performing cytospin preparation, cells were selected based on GFP expression by sorting to enrich cultures on transduced cells. There was a similar erythroblast type distribution among Scrb$^+$ and Sh1$^+$ samples (Figure 31, E).

In theory, none of the populations called negative (Sh1$^-$ and Scrb$^-$) neither Scrb$^+$ should be different because none of them have shRNA silencing RPK. However, Sh1$^-$ was significantly different from Scrb$^-$ (Figure 31, A).

A more detailed study of the transduction efficiency and of the progress of the transduced cells along the differentiation protocol was performed. Approximately, 10-times lower transduction efficiencies were obtained with the sh1 vector than with Scrb vector (n=11). In addition, a EGFP decay (representative experiment: Figure 32, A and B) revealed that the apparent EGFP loss was due to a differential growth capacity of transduced versus non-transduced cells (Figure 32, C and D). Fold increase of GFP$^-$ fractions were a little bit higher than GFP$^+$ (Figure 32, E), which in long term decreased the percentage of transduced cells and limited the number of PKD-like cells available in the cultures.

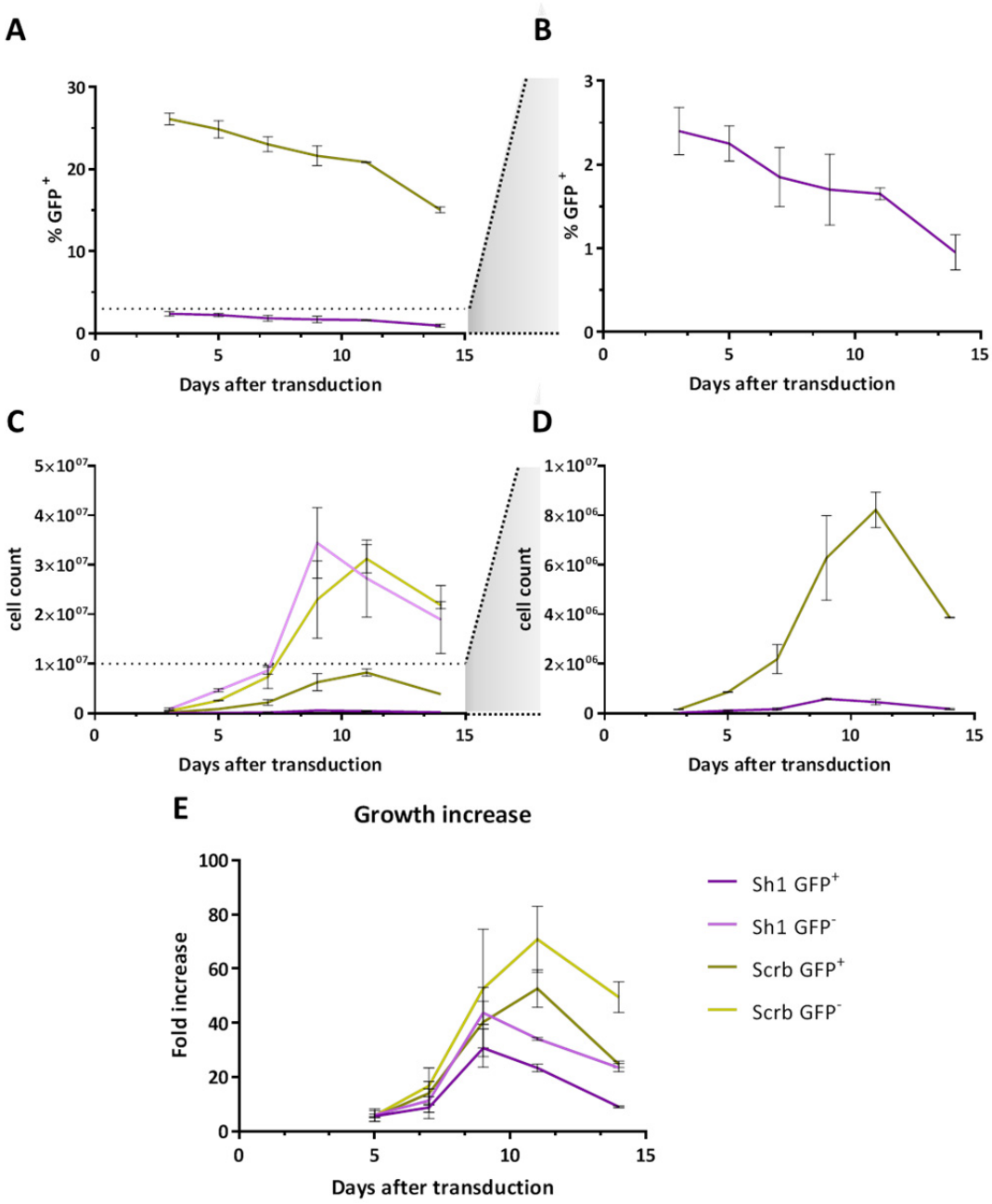


Figure 32. GFP$^+$ decrease observed after shRNA transduction.
Erythroid differentiation began at day 1 post transduction. (**A**) GFP percentage of cells transduced with Sh1 (purple) and Scrb (green) over time. (**B**) Closer view of GFP$^+$ decay of cells transduced with Sh1. (**C**) Total cell count along erythroid differentiation. Dark purple: GFP$^+$ fraction of cells transduced with Sh1; light purple: GFP$^-$ fraction of cells transduced with Sh1; dark green: GFP$^+$ fraction of cells transduced with Scrb; light green: GFP$^-$ fraction of cells transduced with Scrb. (**D**) Closer view of cell count increase of GFP$^+$ fraction. (**E**) Fold increase of cell population over time. Color code used matches with the one described in (C). Due to the different GFP % among experiments, only 2 experiments (out of 11) transduced with the same vector batch are showed in this figure.

Finally, to avoid the problem of the different growth capacity, cells were sorted by GFP and differentiated toward erythroid lineage. Cells silenced with Sh1 decreased PK activity to almost 50%, confirming the inhibition of the RPK mRNA and making $Sh1^+$ modified cells PKD deficient (Figure 33).

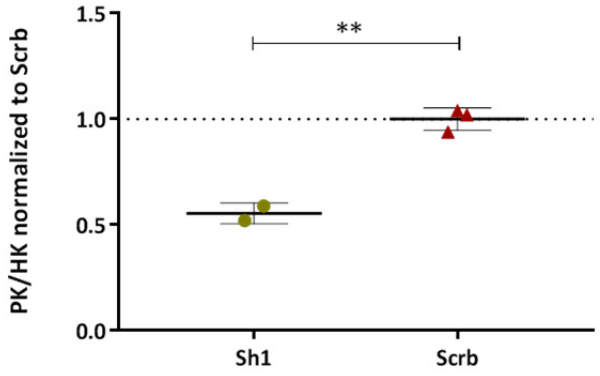


Figure 33. PK/HK activity of cells transduced with Sh1 and Scrb vectors after erythroid differentiation. Green circles: cells transduced with Sh1 vector. Red triangles: cells transduced with Scrb vector. Measures are normalized with respect to Scrb (** P<0.01; Sh1 n=2; Scrb n=3)

2. Generation of PKD-like cells using CRISPR/Cas9 system

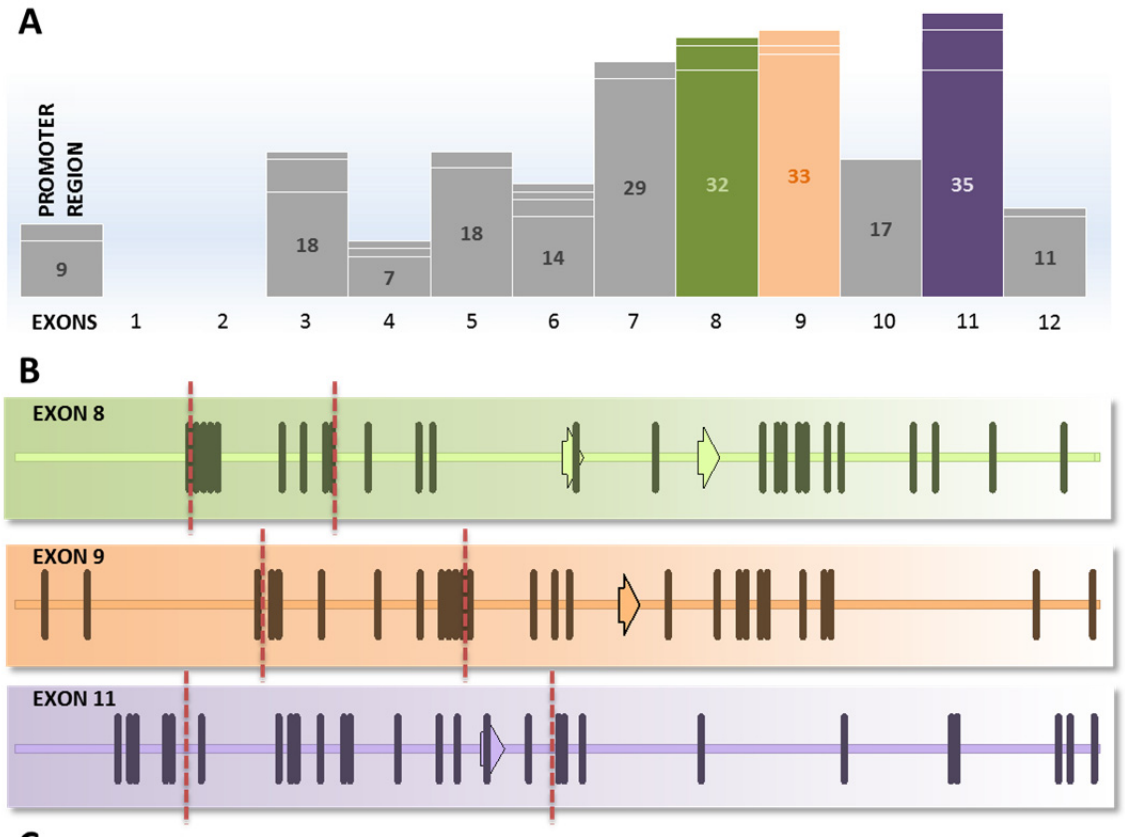
As a second strategy, a genomic knockout of the *PKLR* gene using CRISPR/Cas9 editing tools was addressed. Unlike shRNAs, CRISPR affects directly the genome, modifying the selected gene permanently. The first question to consider was what region of *PKLR* should be edited to generate a good disruption of the gene. Thus, an initial *in silico* study was performed.

2.1 PKD mutations study, *in silico* research of mutation clusters

Among the considered options, a first one to take into account was to directly affect the ATG initial codon region. Consequently, no mRNA would be expected. However, this could be considered an artificial situation since it has not been described in PKD patients. An alternative strategy that could have a more disease modelling outcome was disrupting regions important for enzyme function.

While a precise region-importance relationship has not been stablished, a good indicator of this could be the number of mutations found in PKD patients reported in each exon of the *PKLR* gene.

PKD mutations are differentially distributed along the *PKLR* gene. Among all, exons 8, 9 and 11 are the ones with more than 30 different mutations reported (Figure 34, A).



C

Exon	#	Score	Sequence	Mismatches with respect to coRPK	Strand
8	I	79	CGGCATCATGGTGGCACGGGGGG	4	Sense
8	II	72	GAAATCCTGGAGGTGAGCGACGG	4	Sense
9	I	89	TTACCAAGCCCCGGCCAACGAGG	6	Sense
9	II	64	AGCCCCATCCAGCACAGCATTGG	3	Antisense
11	I	95	TCTGTCTCGGTACCGACCTCGGG	8	Sense
11	II	92	CGCCAGGTCCACTTATGCCGAGG	5	Sense

Figure 34. Guide RNAs design to target *PKLR* gene.
(**A**) Number of mutations reported in each exon. Green (exon 8), orange (exon 9) and purple (exon 11) are the exons with more than 30 different mutations. (**B**) Horizontal lines represent each exon. Vertical lines and arrows are reported mutations, corresponding to one or more than one nucleotide affected, respectively. Vertical red dashed lines represent predicted cut sites of guide RNAs selected. (**C**) Selected guides. Score based on Zhang's Lab (MIT) software. Sequences are written on 5'-3' orientation and include PAM region (red). Mismatches with respect to coRPK are the number of nucleotides different from the therapeutic sequence. Strand column indicates if the guide binds to direct or complementary DNA strand.

Consequently, these were the exons chosen to be the target of the gene edit. Several guides for these three exons were designed using Zhang´s Lab (MIT) software.[125] Among them, only guides with more than 3 mismatches with respect to the coRPK transgene sequence were selected.

Two guides for each exon were selected. Major selection criterion was their proximity to mutation clusters (especially in the case of exons 8 and 9; Figure 34, B). Furthermore, selected-guide's score was above 60 points out of 100 and had at least 3 mismatches with respect to coRPK, meaning that in theory do not recognize the therapeutic vector (Figure 34, C).

2.2 Efficacy of guide RNAs *in vitro*

Six selected guides were electroporated into human lymphoblastoid cells (LCL), to test guide activity. Each electroporated construct contained a guide under U6 promoter and the Cas9 cDNA under the CMV promoter.

After electroporation, cells were amplified *in vitro* for 6 days and then harvested for DNA isolation. PCR of the different exons was performed and then surveyor assay was applied to those amplicons (see *Gene editing evaluation section, Materials and Methods*).

Expected bands were not clearly visible in all the samples of surveyor assay (Figure 35, A). Exon 8 guide I showed a unique band between the two expected ones and guide II did not show any clear pattern. On the contrary, the two guides designed for exon 9 showed the expected band pattern. Finally, in spite of exon 11 guide I surveyor showed some bands that could fit with the gene edit expected, other bands were also present, making these results less clear than those obtained in exon 9. In addition, two different parental bands (about 270 and 300 bp) appeared after editing with exon 11 guide II (Figure 35, A, last column), probably due to unspecific amplification.

The same experiment was performed on $CD34^+$ hematopoietic progenitors purified from healthy cord blood samples. Surveyor assays presented results pretty similar to the ones obtained previously with LCL cells (Figure 35, B). Again, guide I for exon 8 showed a band between the expected sizes and guide II for this exon gave no clear result. On the other hand, the good results obtained for exon 9 were reproduced here on primary cells. No clear

pattern was observed after exon 11 edit. Because of exon 8 guides did not give a clear pattern and exon 11 guides did not produce the expected pattern, they were not selected as candidates for knocking out the *PKLR* gene. Thus, exon 9 guides were the chosen ones for the next experiments.

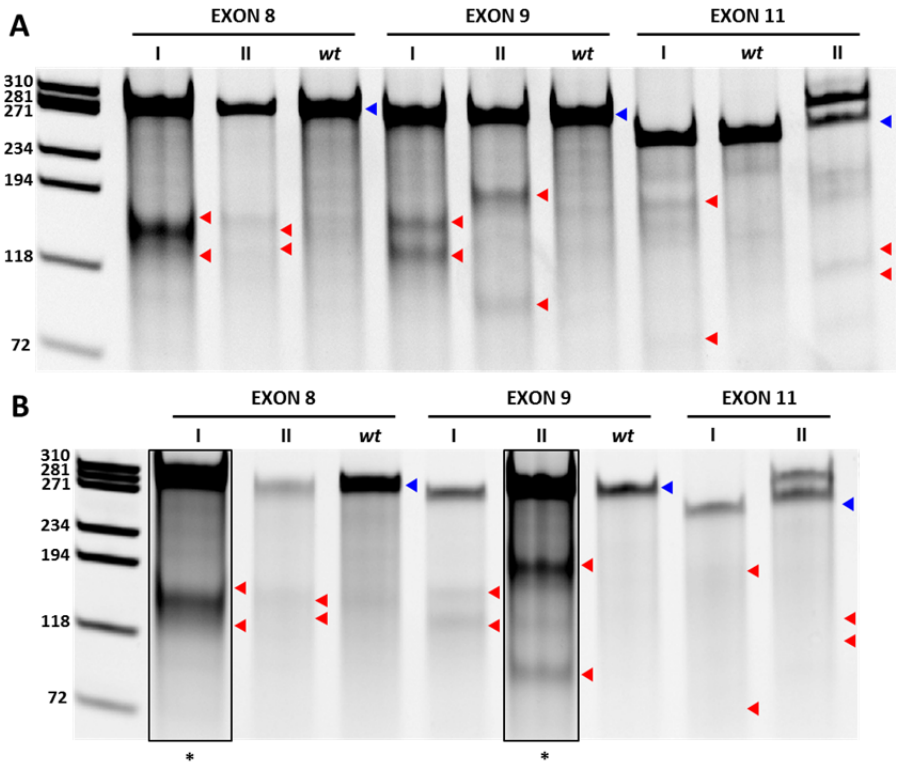


Figure 35. Surveyor assay after gene editing of a cell line and primary cells.
Surveyor assay on LCL cells (**A**) and on primary $CD34^{+}$ cells (**B**) electroporated with 6 different guides, named I and II for each exon. Wt: non electroporated cells. Red triangles: represent bands that correspond with the theoretical sizes expected. Blue triangles: non-cut band that corresponds with the wt PCR amplicon size. Expected bands: Exon 8 (300 bp amplicon) guide I: 166 + 134 bp, guide II: 155 + 145 bp; exon 9 (295 bp amplicon) guide I: 160 + 135 bp, guide II: 191 + 104 bp; exon 11 (259 bp amplicon) guide I: 177+ 82 bp, guide II: 143 + 116 bp. Lanes marked with * correspond to different gels than the rest where result was clearer.

Although exon 9 guides showed good performance cutting DNA, there were still a big proportion of non-edited cells, represented by the parental band on surveyor assays (Figure 35, A and B, 294 bp band). This problem was probably due to the fact that not all the cells were correctly electroporated.

To corroborate this, the same guides were cloned into a plasmid which contained the Cas9 linked to ZsGreen, a reporter gene, by a P2A (motif that allows co-translational "cleavage" of both proteins at equimolar levels). Besides, the plasmid also contained the U6

promoter to express the guide RNAs selected. Hence, just one plasmid was required to introduce the guide RNA, Cas9 and the reporter ZsGreen.

Taking advantage of this fluorescence labeling, GFP⁺ sorted cells would be enriched on the edited population.

Based on this strategy, 293T cells were electroporated with exon 9 guides. Also, a 50/50 mix of both guides was used. 72h after electroporation, cells were sorted and GFP⁺ cells were amplified (Figure 36, A).

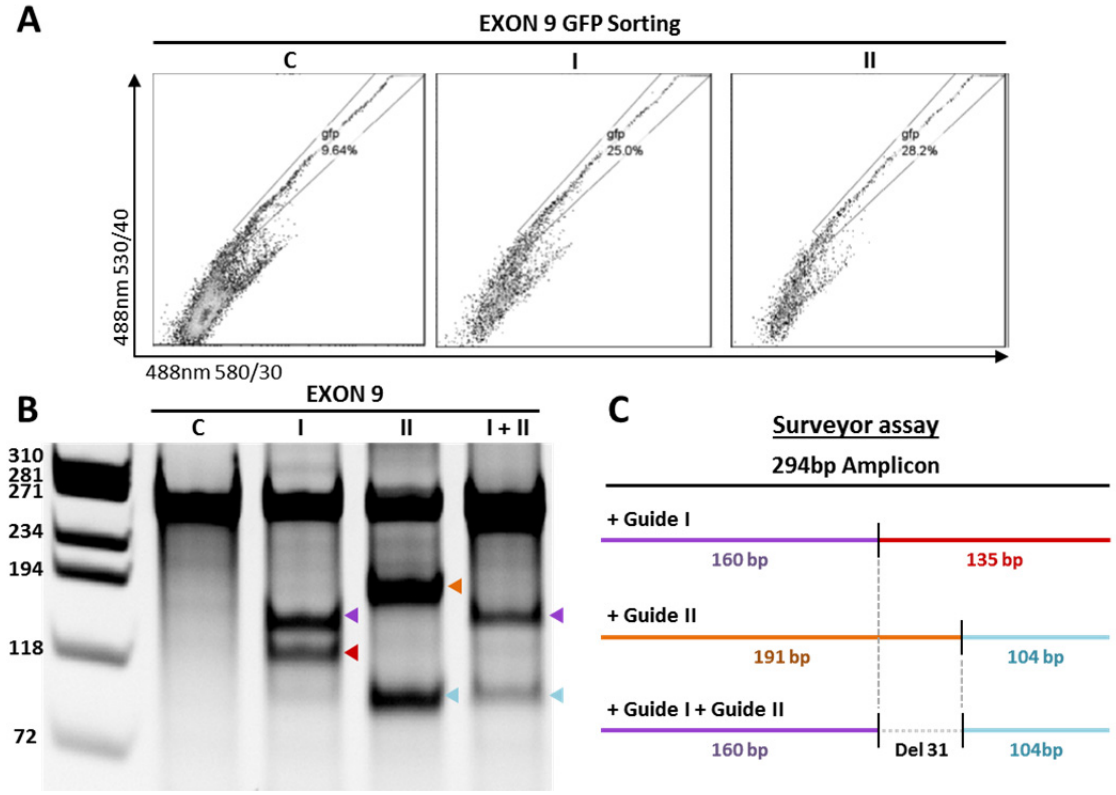


Figure 36. Sorting and analysis of 293T cells edited by guides I, II and I+II.
(**A**) Sorting of GFP$^+$ cells. C, control cells electroporated with a plasmid containing Cas9-P2A-ZsGreen without any guide. I and II, cells electroporated with the guides and Cas9-P2A-ZsGreen. 530/40 vs 580/30 nm dot-plot was used to distinguish GFP fluorescence from auto fluorescence. (**B**) Surveyor assay to check gene editing produced by electroporating guide I, guide II and both together (I+II). Triangles fit with the band sizes expected. (**C**) Scheme of the band patterns that appeared when cells were electroporated with the guides both separately or at the same time.

DNA was extracted and amplification of the region of interest was performed by PCR. Surveyor assay of this PCR amplicons showed that guide I and II presented a good performance, with efficacies of 40% and 59%, respectively, quantified by densitometry. (Figure 36, B). Interestingly, when both guides were electroporated at the same time, a cut

pattern different from the two generated with the guides separately was observed (Figure 36, B, I +II). As can be seen in the scheme (Figure 36, C), this combined pattern fitted with a deletion between both guides. The fact that the other bands (Figure 36, B and C, red and orange bands, 135 and 192 bp, respectively) did not appear on that sample suggested that the proportion of strands cut with a single guide was very low.

2.3 Characterization of guides designed for exon 9

In order to archive a deeper knowledge of the events generated with the different guides and their combination, PCR products of the modified regions were cloned into bacterial plasmids (see *Gene editing evaluation section, Materials and Methods*).

In principle, each amplified bacterial colony contained one event present in the PCR. Individual colonies were sequenced by Sanger method. Twenty five colonies were sequenced for guide I (Figure 37, A), 22 for guide II (Figure 37, B) and 15 for the combination of the two guides (Figure 37, C).

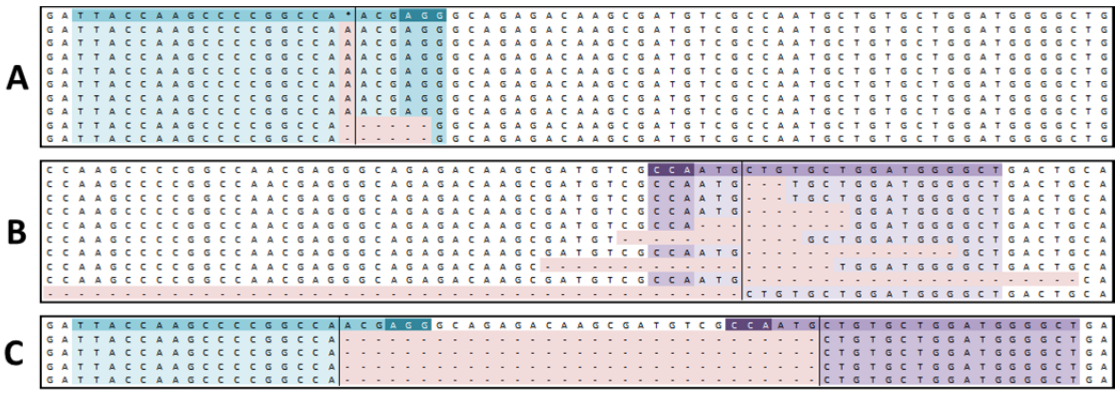


Figure 37. Gene editing events found on 293T cells electroporated with exon 9 guides.
(**A**) 293T cells electroporated with guide I. First row corresponds to wild type sequence, the rest are altered sequences. Dark blue: PAM sequence; light blue: guide sequence; vertical line: expected cleavage region. (**B**) 293T cells electroporated with guide II. Dark purple: PAM sequence; light purple: guide sequence. (**C**) The same cell line electroporated with the two guides at the same time. Color legend is the same as (A) and (B). Red: altered nucleotides, medium dash means deleted nucleotide.

As expected, the use of single guides produced multiple deletions and insertions, around cutting site. Using guide I, 80% of the colonies presented an insertion (A) and a 5 bp deletion in the rest (Figure 37, A). On the other hand, the use of guide II produced different size deletions: 3, 7, 12, 14, 19, 22 and 90 bp around the expected cut site (Figure 37, B).

Interestingly, the only event detected when both guides were used at the same time was a 31 bp deletion. This 31 bp deletion matched with the theoretical precise deletion between nucleotides 3rd and 4th (counting from PAM) of both guides. Despite having sequenced less than 20 colonies, this result, together with the obtained on the surveyor assay (Figure 36, B), suggested that the use of two guides at the same time favors the precise deletion of the DNA between them. This strategy could be a good way to control gene editing events that happen after nuclease activity and the subsequent repair by non-homologous end joining machinery.

Moreover, the precise deletion generated a stop codon 72 nucleotides after the joining point (Figure 38, A). Hence, even in the case of a stable mRNA, approximately one third of the amino acid sequence would be disrupted, including regions essential for the tetramer interaction (Figure 38, B).

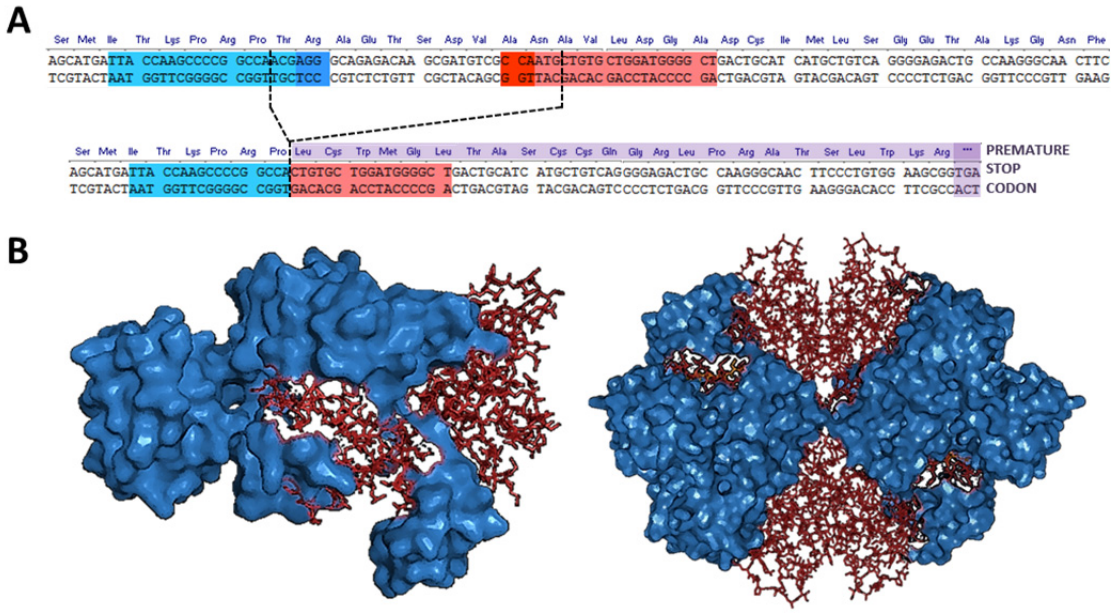


Figure 38. Theoretical gene edit and resulting protein produced by guides I + II at exon 9.
(**A**) Up, Exon 9 *PKLR* gene. Blue highlight is guide I sequence; red highlight is guide II sequence. PAM sequences are marked in darker colors. Dashed line, theoretical cut produced by Cas9. Down, sequence after precise deletion. Purple, new amino acid sequence produced after precise deletion. Dark purple, premature stop codon. (**B**) PK monomer (left) and tetramer (right); blue surface represent regions that conserve wt amino acid sequence, red sticks are amino acid changed after deletion. (Model based on PDB: 2VGB [18]).

To further explore the possibilities of controlling gene edit with the CRISPR/Cas9 system using two guides at the same time, 5 new guide combinations were generated.

2.4 Analysis of the deletion efficacy using two guides simultaneously

Based on the idea of generating controlled events deleting regions between guides, 5 different additional combinations were tested. These new combinations, together with the first one (31bp), hereinafter called combos, were meant to generate deletions of different sizes, of approximately 10, 30, 60, 120, 240 and 480 bp (Table 10). All the combos shared guide I, so every deletion was performed starting from *PKLR* exon 9 (Figure 39, A).

Table 10. Deletion distances using 6 guide combinations.

Name	Combo 1	Combo 2	Combo 3	Combo 4	Combo 5	Combo 6
Planned distance	10	30	60	120	240	480
Real Distance	8	31	63	119	239	490

Guides were cloned in the same construct containing Cas9-P2A-ZsGreen to take advantage of the sorting strategy. Pairs of guides were electroporated into K562 cells. Thanks to the endogenous RPK expression of this cell line, the efficacy of gene editing could be also corroborated checking the decrease on protein production. Again, GFP^+ cells were sorted to enrich edited population. DNA was isolated to verify the deletion and protein was extracted to evaluate the effect at protein level.

Combos 2 to 6 presented smaller PCR amplicons compared to control one that perfectly corresponded with the expected deletions (Figure 39, B). Alternatively, combo 1 deletion was too small to be observed due to the resolution of the gel used. In any case, GFP selection was confirmed as a very effective selection method since wt band was not visible in any of the deletions.

Western blot analysis showed that RPK protein was almost completely eliminated in all combos (Figure 39, C) meaning that the edited region was crucial for protein generation.

The results observed were clear in terms of PCR amplicon sizes and in terms of protein production. Nevertheless, to deeply analyze how precise was the removal of the DNA sequence between the guides, the amplicons were deep sequenced by next generation sequencing.

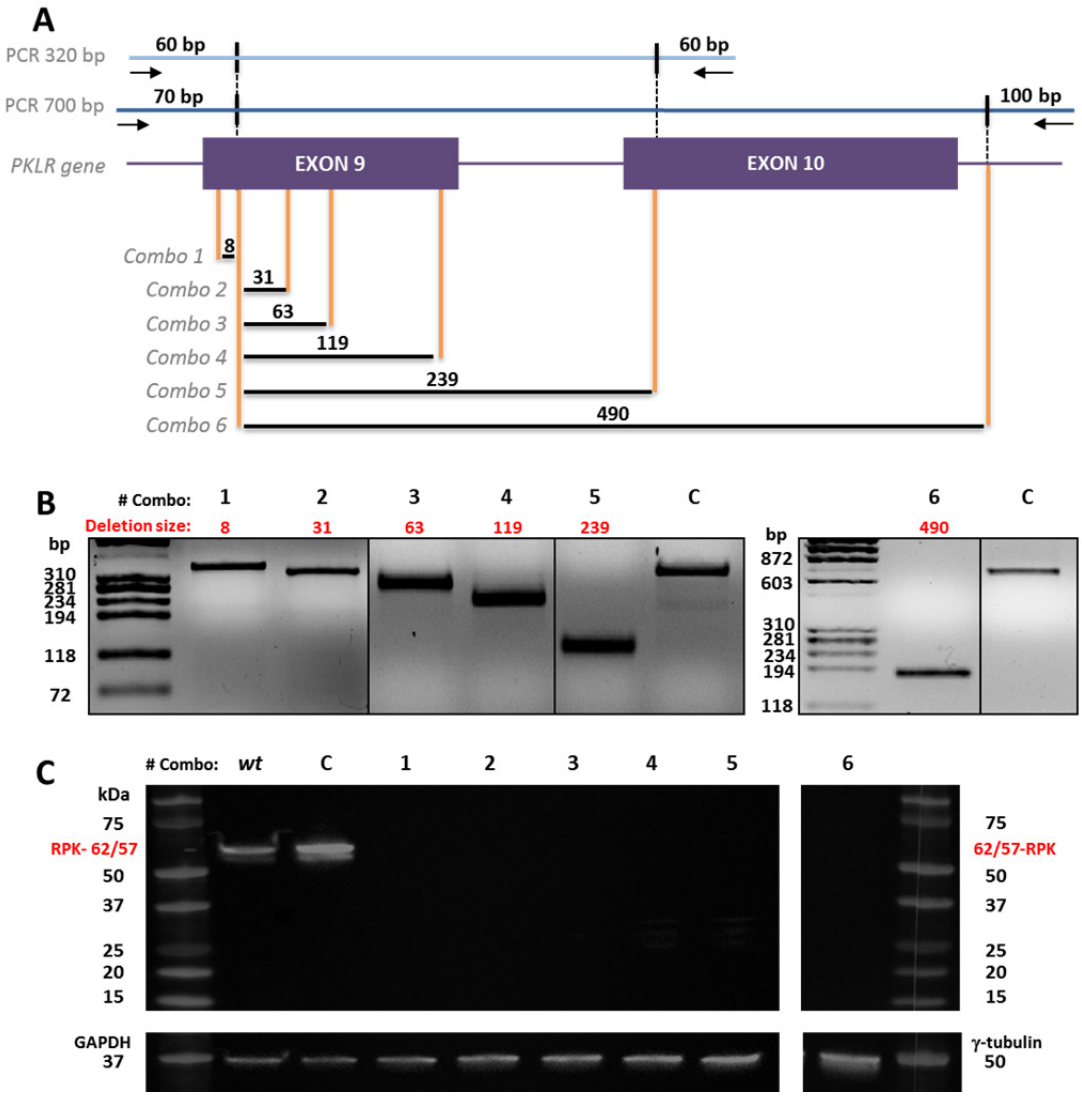


Figure 39. DNA and protein analysis of K562 cells electroporated with 6 guide pairs.
(A) Scheme of 6 combos tested. Light blue: PCR amplicon used for the analysis of combos 1-5; Dark blue: PCR amplicon used for combo 6 analysis. In purple: *PKLR* gene, exons are represented as boxes and introns as horizontal lines. Orange vertical lines: position of the guides used. Numbers above black horizontal lines are the size of the deletion generated in each combo. (**B**) PCR amplicons of exon 9 from different combos. Up: combo names. C: control sample, K562 cells electroporated with Cas9-P2A-ZsGreen without guides and sorted. Red: size of the precise deletion. Combos 1 to 5 were evaluated performing a 320 bp PCR meanwhile combo 6 were checked by a 700 bp PCR. (**C**) Western blot of RPK from combos. Wt: non-electroporated K562 cells. C: control sample, K562 cells electroporated with Cas9-P2A-ZsGreen without guides and sorted. Red: sizes of RPK protein, complete (62 kDa) or after physiological N-terminal excision (57 kDa). GAPDH and γ-tubulin were used as loading control.

The prevalence of the precise deletion as the most represented event was maintained along all tested distances except in the 8 bp deletion (Figure 40). A deeper study of the events generated trying to produce the 8 bp deletion demonstrated that all of them were caused by guide I action alone (Figure 40, A, guide I is represented in blue). In fact, double

strand break generated by guide I also altered the other guide sequence (Figure 40, A, orange), preventing guide 8 joining and, hence, not making possible the precise deletion.

The combination of guide I and guide II, also named combo 2, showed two unique events. Precise deletion (the 31 bp removal, Figure 38, A) already detected by Sanger sequencing (Figure 37, C), and a 32 bp deletion. Precise deletion represented 87% of events recorded. The other event registered, a 32 bp deletion, shared the 31 bp deletion with the most represented minus an adenosine located 5' of the guide I cut (Figure 40, B).

Combos 2 to 6 showed the precise deletion as the most frequent event (69-96%) produced (Figure 40, B-F). In combos 3 to 6 (63-490bp deletions) additional indels (insertion/deletion events) near the precise deletion (Figure 40, B-F, *IND*) were found. Furthermore, some indels were found produced by guides separately in combos 3 to 6 (Figure 40, *SGI*, above wt and guides sequences).

Unfortunately, this strategy of double mapping does not permit the comparison of the frequencies of the events produced by the two guides together (Precise deletion and *IND*) and the frequencies of guides that acted individually (*SGI*). However, this could be indirectly addressed comparing the proportion of different sizes PCR products. Thus, SGI had to be barely represented as long as wild type PCR band was not visible in the agarose gels (Figure 39, B).

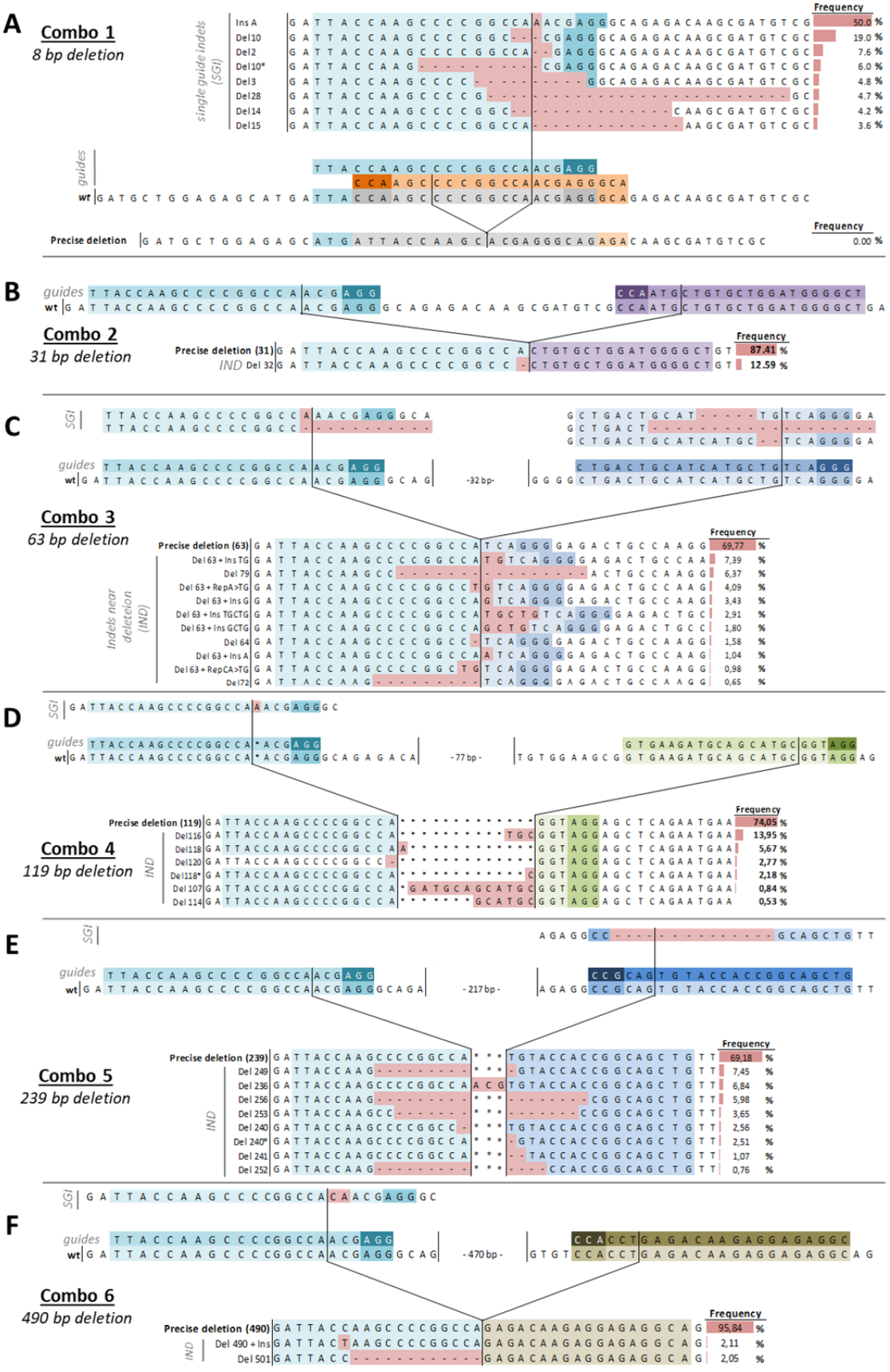


Figure 40. Variants and its frequency of combos 1-6 analyzed by NGS.
Within each combo scheme there are three differentiated parts. Up: indels generated by guides individually, named single guides indels (*SGI*). Middle: wt sequence and the guides used for each combo.

Down: Precise deletion sequence and its frequency. Bottom: indels near precise deletion and its frequency (*IND*). NGS was performed for two independent experiments per each deletion distance.

The use of guide pairs was an effective strategy to control gen editing events in distances from 31 to 490 bp. Moreover, the proportion of precise deletion was not affected by the progressive increase of the distance between guides (Figure 41).

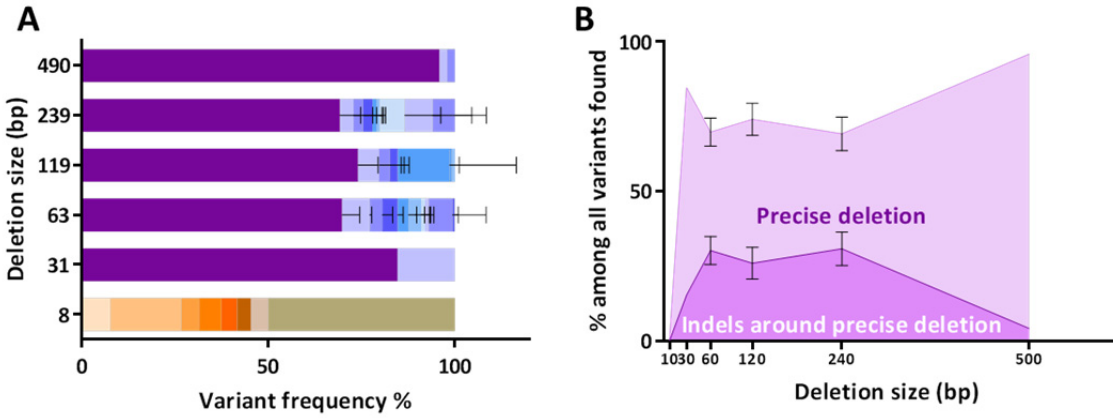


Figure 41. Precise deletion compared with the distance removed.
(A) Frequency of the variants found in the deletions. Purple bars: Precise deletion. Different type of blue bars: Indels near precise deletion (*IND*). Different type of orange bars: indels generated by just one guide (*SGI*). **(B)** Proportion of precise deletion and indels near precise deletion with respect of the distance between the guides used.

2.5 Generating PKD-like cells from hematopoietic progenitors

Having evaluated the possible outcomes of gene editing on K562 cells, the next step was to apply the strategy to primary cells. It was already demonstrated that any of the guide combinations tested were able to disrupt RPK production with a good performance. The combination chosen was the combo 2, the guides that generated a 31 bp deletion in exon 9 (guide I and guide II, hereafter named I+II) because the events generated were tightly controlled: only two type of deletions (31 and 32 bp) according to NGS data (Figure 40, B). Furthermore, both deletions altered the amino acid sequence, generating premature stop codons after 24 amino acids in the case of 31 bp deletion and after 5 amino acids in 32 bp deletion case (Figure 42, A).

Hematopoietic progenitor cells ($CD34^+$) obtained from healthy cord blood samples were electroporated with guides I + II at the same time. Twenty four hours after electroporation, cells were sorted based on GFP expression. Then, a two week erythroid

differentiation protocol was performed. Cells were analyzed in terms of DNA editing efficacy, immunophenotype, cytomorphology and pyruvate kinase activity.

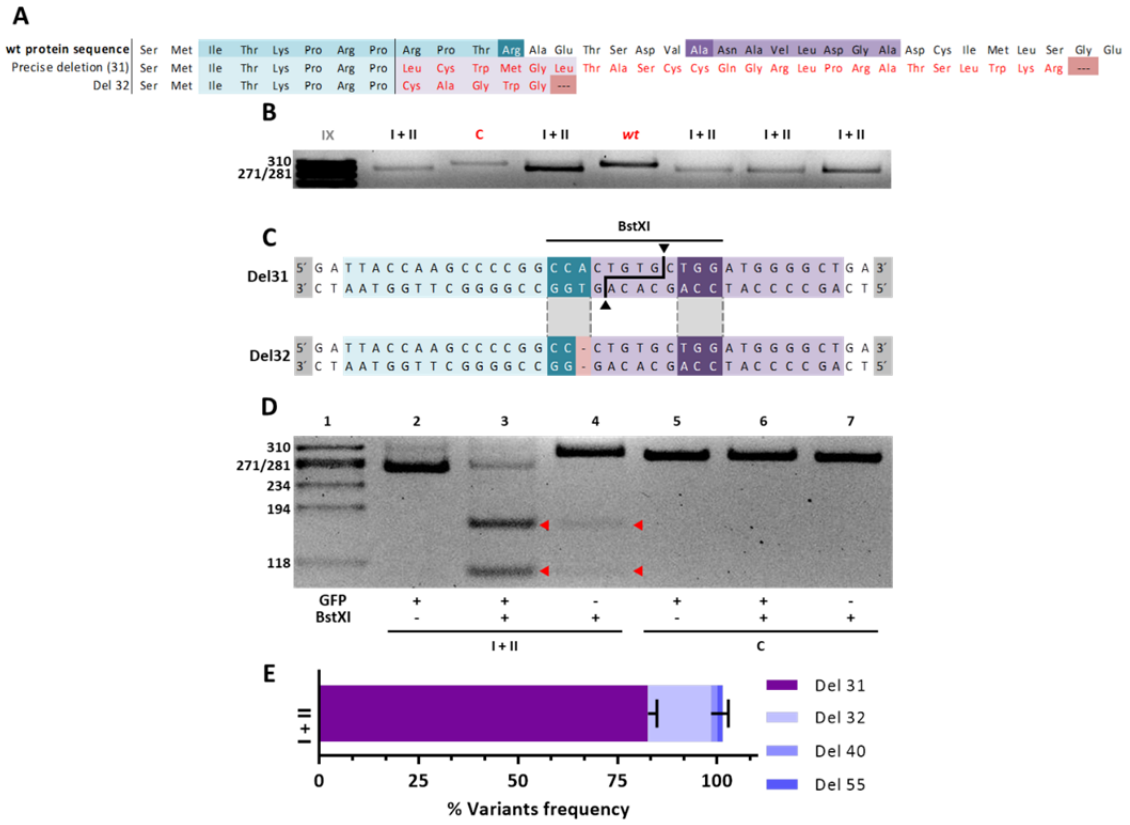


Figure 42. gDNA study of primary edited cells after erythroid differentiation.
(**A**) Amino acid sequence resultant after 31 and 32 bp deletions. Red amino acids are the alerted ones with respect to the wt. Both deletions generate premature stop codons represented by "…" on red background. (**B**) Exon 9 PCR amplicons from different experiments. I + II: edited cells. C: control cells electroporated with Cas9- P2A-ZsGreen without guides. I + II and C are the GFP^+ fraction selected by sorting. Wt: non-electroporated cells. (**C**) Sequences of the two main deletions produced after gene edit. Blue: region covered by guide I. Purple: region covered by guide II. Sequence from dark-blue to dark-purple: BstXI recognition site. Light red: differential nucleotide, not present in 32 bp deletion, that is required for BstXI recognition. (**D**) Lane 1, molecular weight reference. 2-4 lanes: hematopoietic progenitor cells differentiated *in vitro* toward erythroid lineage and edited with guides I + II. 5-7 lanes: control cells electroporated with Cas9-P2A-ZsGreen without guides. GFP, + or – fractions separated by sorting. + or – BstXI, treated or not with the enzyme. Red triangles, DNA cut after enzyme digestion. (**E**) Next generation sequencing results from two edited DNA samples, colors represent each of the deletions found named Del 31, Del 32, Del 40 and Del 55.

DNA edit was the first checkpoint to validate that the cells were edited. The use of these two guides generates a 31/32 bp deletion easy to detect by PCR due to decreased band size with respect to wt sequence. This was consistently observed in every independent experiment carried out. Five of them are presented on Figure 42, B.

The precise deletion generated a new restriction site in the DNA. Thus, a fast way to corroborate the presence of this 31 bp deletion was based on the presence of a new BstXI site. It appears only in the case of the precise deletion (31 bp) but not in the other deletion (32 bp; Figure 42, C). After digesting the PCR product with this restriction enzyme, DNA cut pattern appeared in cells electroporated with guides I + II (Figure 42, D, lane 3). It was also possible to observe a faint cut on the GFP^- fraction, meaning that maybe some edited cells were not GFP at sorting step (Figure 42, D, lane 4).

Two edited samples, differentiated to erythroid lineage, were also analyzed by NGS. Interestingly, on primary cells the same two deletions appeared almost in the same proportion as in K562 cell line: about 83% were precise deletion (31 bp) and 16% were Del 32. In one of the samples, two other deletions were also found (Del 40 and Del 55) in a low frequency, 1.6 and 1%, respectively (Figure 42, E).

The next step was to find out if this knockout of *PKLR* gene affected *in vitro* maturation of the hematopoietic progenitors to mature erythroblast forms. As it has been commented previously, hematopoietic progenitors, after being electroporated with guides I + II and sorted, were differentiated *in vitro* based on the 3 weeks protocol explained (n=6; see *Erythroid differentiation protocol, Materials and Methods*).

Cytometry analysis was performed to measure the proportion of cells within the transition from $CD36^+/CD45^{mid}$ to $CD36^-/CD45^-$ (all of them $CD235a^+$), which represents the proportion of cells maturing along the erythroid lineage (Figure 43). Gating strategy was the same used on shRNA section without the first gating based on GFP since at the end of differentiation protocol all cells were GFP negative (Figure 30, B and C).

In parallel, hematopoietic progenitors from bone marrow sample of a PKD patient (c.359C>T, exon 4/ c.1168G>A, exon 9) was differentiated *in vitro* to validate that the model was representing correctly the disease.

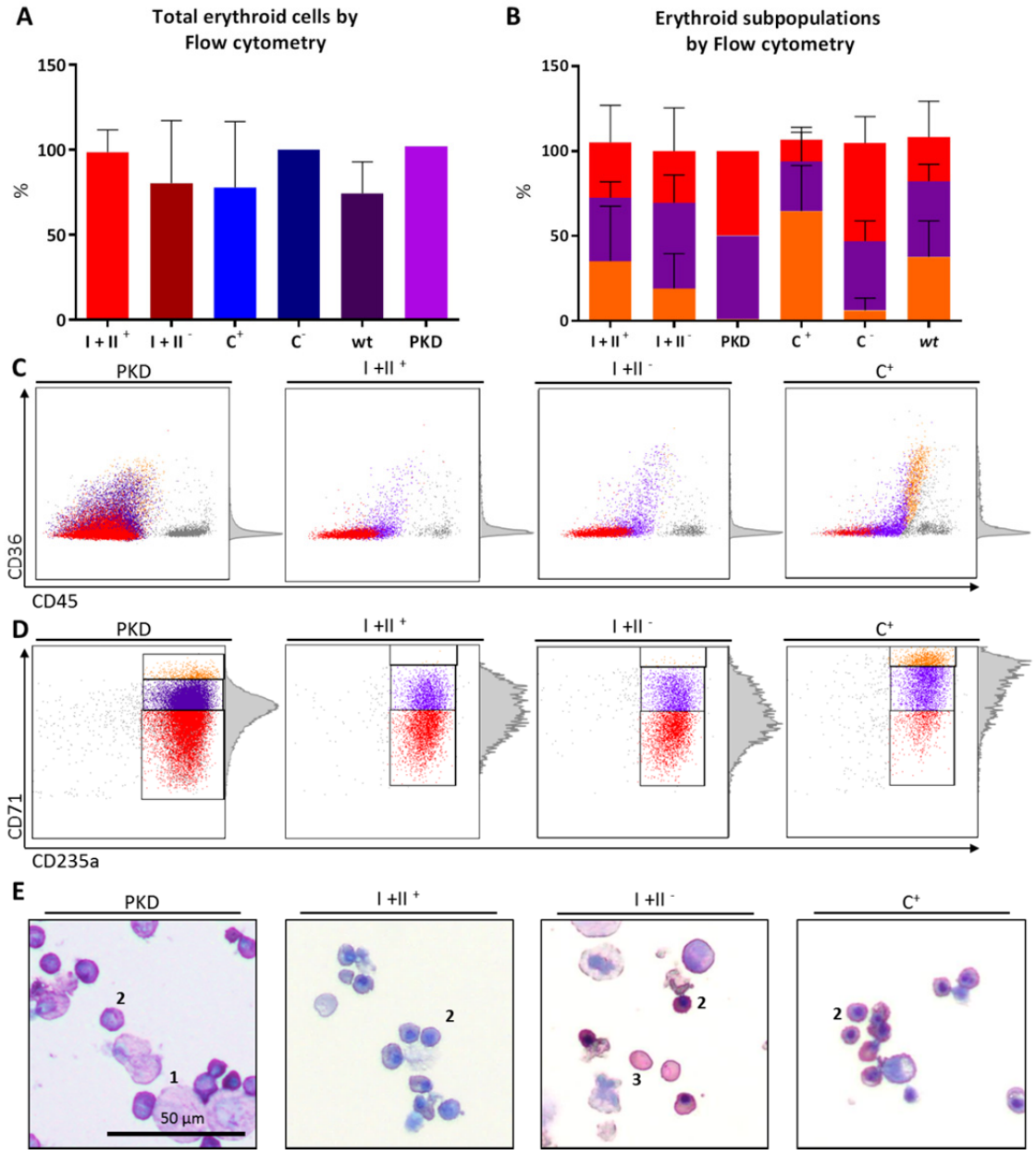


Figure 43. Hematopoietic cells edited and differentiated towards erythroid lineage.
PKD: cells differentiated from patient bone marrow sample. I + II: cells electroporated with the two guides. C: control cells electroporated with Cas9-P2A-ZsGreen without guides. + or -: GFP positive or negative fraction separated by sorting. Wt: non-electroporated/sorted cells. (**A** and **C**) Cytometry results for CD36/CD45 staining. Gray: histogram that represents the cell distribution along CD36 staining. (**B** and **D**) Cytometry results for CD71/CD235a staining. Gray: histogram that represents the cell distribution along CD71 staining (n=6). (**E**) Cytospin after 14 days of erythroid differentiation. 1: polychromatophilic erythroblast; 2: orthochromatic erythroblast; 3: reticulocyte.

As in the shRNA study, results were normalized with respect to the control, C^- (GFP^- cells electroporated with Cas9-P2A-ZsGreen without guides that is the sample more similar to the wild type). There were not statistical differences between edited cells ($I + II^+$) and any control, such as $I + II^-$ (GFP^- fraction of cells electroporated with the guides), C^+ and

C^- (GFP^+ and GFP^- fractions of cells electroporated with the plasmid without guides). Also, there were no differences with respect to the non-electroporated cells (wt) and neither with the bone marrow $CD34^+$ PKD cells differentiated from the patient bone marrow sample (Figure 43, A and C).

Equally, the cell distribution from more immature II ($CD36^+/CD45^{mid}$) to more mature IV ($CD36^-/CD45^-$) cells did not show statistical differences. Despite the mean of subpopulation II of C^+ was higher than the rest; it also had a huge dispersion meaning that there were no consistent differences (Figure 43, B and D).

That lack of differences was also observed at the microscope. Both deficient (I + II^+ and PKD) and control cells (I + II^- and C) reached the stage in which most of the cells were orthochromatic erythroblasts (Figure 43, E, 2)

Finally, PK/HK ratio was analyzed in *in vitro* differentiated cells to check if the RPK activity had been impaired by the gene editing protocol (Figure 44).

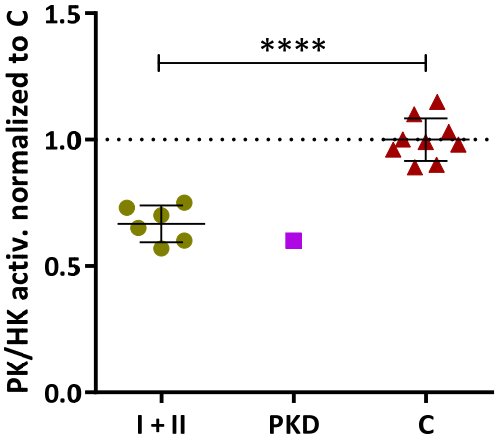


Figure 44. PK/HK activity of edited cells after erythroid differentiation.
Green circles: GFP^+ fraction cells electroporated with the two guides from different experiments. Purple square: PKD cells differentiated from patient bone marrow sample. Red triangles: mix of different control cells electroporated and sorted (GFP^- of cell cultures electroporated with the two guides; GFP^+ and GFP^- populations of cells electroporated with Cas9-P2A-ZsGreen without guides). Measures are normalized with respect of PK/HK of control samples, which were consider =1 (**** P<0.0001; n=6)

Edited cells, I + II, reached PK/HK values comparable with the obtained in the *in vitro* differentiated PKD patient sample and significantly lower than the values obtained from control wt differentiated cells.

On the whole, it has been demonstrated that I + II cells had the expected DNA modification (Figure 42), that this modification impaired RPK production (Figure 39) and, despite cells were capable of differentiating along erythroid lineage (Figure 43), their PK/HK activity was as much compromised as on patient cells (Figure 44). Hence, I + II cells have demonstrated being a useful tool as new *in vitro* model to study the disease on human cells.

Discussion

Discussion

PKD patient features

Analysis of PKD patient peripheral blood samples corroborated the main features described for the disease, such as RBC, hemoglobin and hematocrit decrease and MCV increase of patient´s RBC due to the egressing of immature RBC from bone marrow to compensate the anemia.[48]

The use of a system such as ROC curves, based on the representation of true positive rate against false positive rate at different thresholds (understanding positive as PKD patient and negative as healthy donor), has been useful to set predictive values of hematological parameters (hemoglobin, RBC count, MCV, etc.) that permitted to define samples as PKD. The few PKD-relatives analyzed, despite they are not affected by the disease, showed a profile closer to PKD patients than to healthy donors. Nevertheless, relatives' values remain within the healthy side of the thresholds set based on ROC curves, predicting them correctly as healthy. These thresholds are meant to distinguish PKD patients from their relatives and healthy donors but are not values for diagnosis. However, clinical values of hemoglobin levels considered anemic are 11-11.9 g/dL in women older than 15 years old and 10-12.9 g/dL in men older than 15 years old. The average of the clinical criteria (11.45 g/dL) was below ROC curve threshold set for hemoglobin in PKD patients (12.35 g/dL).[130]

ROC curves were also performed on flow cytometry panels for the characterization of the anemia compensatory mechanism. Before an anemic situation, bone marrow releases erythroid progenitors ($CD71^+$ $CD235a^+$) and reticulocytes trying to compensate deficient RBC function. This was observed on PKD patients but not in their relatives, meaning that having just one allele mutated does not decrease hematological traits enough to push bone marrow to an anemia response. In this case, ROC curves thresholds help to identify if the patient is anemic enough to trigger the bone marrow response.

However, the decrease of all these hematological parameters is common to other anemias, so they cannot be considered the only diagnostic criteria, but they are very useful to follow-up the progression of the disease.

Another altered value found in PKD samples was monocyte counts when they were estimated using the hematological counter. In spite of the dispersion, mean PKD values were statistical lower than control ones. It must be taken into account that hematological counter analyzes leukocyte subsets as a relative proportion of three population (monocytes, granulocytes and lymphocytes), meaning that the registered monocyte decrease can be because of a real reduction of this population or because of an increase in the absolute number of granulocytes and lymphocytes. This population was also evaluated by flow cytometry through the analysis of the CD14 surface marker and indeed, PKD $CD14^+$ percentage was slightly lower than healthy values (5±2 vs. 7±3, mean±SD). Nonetheless, values were so disperse that they did not give a statistical relevance to this difference. The disparity of the two methods and the fact that monocytes decrease has not been described before in bibliography, make us think that the changes were probably the result of artifacts of the procedure used. Additional patients will be analyzed to clearly define this potential decrease of monocytes in PKD patients. Apart from monocytes, no alterations were found in immune system populations, matching with values described in the literature.[48,51]

One of the hypotheses posed was that knowing the fact that the bone marrow reacts to anemia by increasing the fraction of reticulocytes in peripheral blood, it could also affect to more immature hematopoietic populations. In other words, if PKD anemia triggers an egress of stem hematopoietic progenitors that would be reflected in an increase of circulating $CD34^+$. Percentages of $CD34^+$ cells in peripheral blood samples of PKD patients were marginally higher than in healthy donors. Something similar happened in the CFU assay, with a minor increase of the colonies originated in the patient´s samples, but not statistically significant. Thus, although a clear tendency is observed, the effect is very mild. A possible explanation could be that only in the most severe cases PKD affects the most primitive subsets of the hematopoietic lineage, something that has been described in a few cases of PKD whose alterations led to misdiagnosis, being confused with congenital dyserythropoietic anemia.[131] Again, higher number of healthy donors and patients, including severe cases, should be analyzed to clearly define this potential increase. Moreover, this increase of hematopoietic progenitors in peripheral blood could be also inducing the generation of new niches of extramedullary erythropoiesis since splenic and hepatic erythropoiesis has been described in few PKD cases. Most of these cases were limited to Amish communities, therefore, it would be interesting to evaluate the level of

primitive hematopoietic progenitors ($CD34^+$ cells or CFU) in these population to study if there is a relationship between the levels of hematopoietic progenitors in the peripheral blood and levels of extramedullary hematopoiesis..[51,132]

In any case, no impairments were demonstrated in PKD $CD34^+$ capacity to generate CFUs. This has special relevance thinking in the eventual clinical trial protocol in which coRPK vector is going to be used. The planned clinical protocol is based on autologous transplant of *ex vivo* $CD34^+$ corrected cells. PKD hematopoietic progenitor collection would be performed by bone marrow mobilization, thus, robustness of hematopoietic progenitors would facilitate this step. Moreover, if finally this slight increase of $CD34^+$ cells in peripheral blood is demonstrated, it could make $CD34^+$ collection phase even easier.

Regarding the mutations-severity association, a recent PKD natural history study pointed out that double non-missense (NM/NM) mutation could indicate a more severe phenotype, while double missense (M/M) predicts a better prognosis.[51] Data collected from the patient´s analysis followed the same tendency. However, the number of patients per group was not enough to evaluate these data statistically. Finally, no severity-gender association was found in the analyzed patients.

Altogether, these data have allowed us to carry out a complete characterization of peripheral blood of PKD patients and to compare it with their own relatives and healthy donors.

Towards gene therapy clinical trial for PKD

As commented before, the set-up of the gene therapy clinical protocol and clinical trial for PKD requires the definition of important details, such as the amount of viral particles needed to cure the disease. This number depends directly on viral vector efficacy and engraftment capacity of the transduced hematopoietic progenitors and stem cells, and would be reflected by the number of transduced cells required to restore a healthy erythroid phenotype. This final outcome can be followed-up by measuring the VCNs present in the peripheral blood cells of the treated patients. Previous studies have pointed out the therapeutic threshold above 25% of corrected cells or above 1 VCN [79,124]. We wanted to define the minimum number of coRPK vector copies per cell required to restore a healthy

phenotype in the PKD mouse model available in our laboratory. This was crucial to make feasible gene therapy for PKD.

ROC curves were performed using hematological parameters of C57BL/6 mice as healthy reference and AcB55 mice as PKD. The criteria for the election of the thresholds generated with these ROC curves were to maximize the sensitivity (thresholds that includes the highest number of true positives) and also the specificity (thresholds that includes the lowest number of false positives). Setting these thresholds was useful to distinguish healthy and deficient mice due to that, even in these control mice, most of the parameters (erythrocyte blood count, hemoglobin, etc.) had overlapped healthy/deficient values. For other features, such as reticulocyte percentage and spleen weight, in which healthy/deficient values were clearly separated, standard deviation range was used as reference.

Regarding the use of standard deviation of reticulocyte percentages as reference, it must be taken into account that all experiments performed in mice were based on conditioning and transplantation of corrected hematopoietic progenitors. After transplant, all mice had to reconstitute all hematopoiesis and showed an increase in reticulocyte count that tended to stabilize, including mice transplanted with healthy hematopoietic progenitors (unpublished data from the group). Thus, reticulocyte values of healthy controls (2-4%) were lower than the recently transplanted mice (10-20%), so ROC curve threshold for reticulocyte percentage based on non-transplanted healthy mice would considered deficient even mice transplanted with healthy control cells.

The first experiment performed using two coRPK transduction rounds achieved a complete curation of the mice, restoring healthy values of PK activity at day 40. Moreover, after a year of follow-up there were no evidences of anemia, reaching almost healthy values of reticulocyte percentages. Indeed, VCNs obtained were at least 10 times higher than the therapeutic threshold stablished later on. In this experiment, VCNs obtained were very variable, ranging from 2 to 6, and not related to the MOI used. This variability could be due to the conditions used in the transduction protocol, like the use of two transduction rounds, or to the transplantation of low number of stem cells transduced. Nevertheless, all mice were above the therapeutic threshold intended.

The use of only one transduction round and double number of cells transplanted per mouse ($1x10^6$ instead $0.5x10^6$) showed less variability in VCN with respect to MOIs used.

Correlation of VCN with respect to reticulocyte percentage, the reference parameter to follow-up the progression of the disease in mice, permitted to stablish a narrow window between 0.2 and 0.3 coRPK copies per cell above which mice were able to overcome the disease.

Similar results were obtained in other experimental series performed in our laboratory where a different approach was addressed (unpublished data). Instead of transplanting coRPK corrected cells, healthy syngeneic hematopoietic cells were used. Cells were transplanted in different proportions of wild type and PKD (from 5 to 100%). Thirty percent of healthy donor cells were needed to have a therapeutic effect and to correct the disease. Consistently, data obtained in both experimental series demonstrated that a minimum of 30% healthy or corrected cells was sufficient to compensate the disease, setting the threshold above which our gene therapy program should be successful.

These VCN values are clearly lower when compared with the copies per cell achieved in clinical trials developed by other groups that work in hematopoietic diseases. For instance, in different studies conducted to treat β-thalassemia between 0.7-2.1 copies per cell were required and close to 1 copy per cell in another trial conducted for Wiskott-Aldrich syndrome.[133–136]

On the other hand, the therapeutic coRPK vector has demonstrated to be able to reach 1 VCN per cell at MOI 50 in the mouse model. Thus, optimized conditions able to reach 1 VCN in the human hematopoietic graft would ensure the therapeutic threshold and the feasibility of the gene therapy clinical trial for the treatment of PKD patients.

Gene therapy has demonstrated being a feasible treatment for other hematological diseases[133–137] and we are the first group trying to apply it to PKD. This will mean a new perspective for the most severe patients that are transfusion-dependent even after splenectomy and whose only current curative alternative is allogenic bone marrow transplantation.

Human PKD modelling

As it has been previously commented, the study of gene therapy and gene editing approaches to treat PKD had a main drawback, the lack of hematopoietic progenitors from patients. The target cell type for these therapies would be obtained from PKD bone marrow samples, but this sampling is too invasive and it is not part of the follow-up of PKD patients. To overcome this lack of PKD progenitor cells, two molecular strategies were carried out to generate PKD-like cells modifying healthy $CD34^+$ progenitors.

Regarding the generation of PKD-like cells by specific shRNAs, it was demonstrated that the three vectors tested were effective decreasing RPK protein levels in K562 cell line, especially Sh1. More importantly, the activity of the RPK protein, measured as the ratio PK/HK at the end of the differentiation protocol, was reduced up to 50% when compared with non-interfered cells, demonstrating the efficacy of the Sh1 viral vector to induce a PKD-like phenotype. However, surface antigen expression profile of the different erythroblast populations along the *in vitro* differentiation protocol did not show specific differences in the RPK interfered cells ($Sh1^+$) when compared with either the differentiated cells after transduction with the Scrb vector or the $Sh1^-$ cells, the cells generated in the same culture coming from non-transduced progenitors. In the same way, the *in vitro* erythroid differentiation of $CD34^+$ cells of a patient sample (see *Generation of PKD-like cells using CRISPR/Cas9* section), as well as the PKD-like cells generated with the CRISPR/Cas9 system showed that pyruvate kinase deficient cells did not present differences in the antigen expression profile along *in vitro* differentiation in comparison with normal cells. These results demonstrate that the lack of RPK does not impaired the *in vitro* differentiation profile and that the only way to study RPK inhibition is by analyzing the functional activity of the RPK protein.

Two unexpected limitations were found regarding inhibition of the RPK expression by shRNAs transduced with lentiviral vectors that compromised the use of this approach as model. The first one was the low transduction efficacy of Sh1 vector, which varied from 2 to 20% depending of the experiment and of differential vector batch efficacy (5 different batches were used). The second unexpected limitation was the apparent decrease of the EGFP positive population over time on both Scrb and Sh1 transduced cells, resulting in a differential growing rate between transduced and non-transduced cells. The use of shRNA

lentiviral vectors seemed to slow down cell growth and differentiation towards erythroid lineage. The fact that these differences were also found in cells transduced with the Scrb vector indicates that this impairment was not due to a specific silencing of *PKLR* gene. Furthermore, the RPK protein starts to be crucial at polychromatophilic/orthochromatic stages, which only occurs at the end of erythroid differentiation protocol, meaning that the effect could be more related to the viral vector used. This unspecific growth arrest has not been reported with shRNAs under H1 promoter (like the used in this work) but it was observed with shRNAs under U6 promoter.[138] A possible explanation to this effect is an innate interferon-mediated response probably triggered by the presence of dsRNAs. Other authors showed that this response could be dependent on slight differences in the shRNA sequences, making some of them more prone to initiate an immune response.[139]

It is also important to remark that cells transduced with the shRNAs and with the therapeutic vector (coRPK) were able to produce the transgenic protein, meaning that any developed shRNA recognized coRPK mRNA, as expected because all the shRNAs had 3 mismatches with respect to the codon optimized sequence, or because the level of expression of the transgene was too high to be silenced by Sh1.

Altogether, despite the growth issues found along *in vitro* differentiation protocol, the Sh1 lentiviral vector showed a high and very specific capacity to silence *RPK* gene expression. Unfortunately, the low number of PKD-like cells generated in these cultures limited its potential use as a human source of PKD-like hematopoietic progenitors.

As a second strategy, we developed a genome editing approach to knock-out the *PKLR* gene. This strategy was possible thanks to the development of highly efficient engineered nucleases that induce double strand breaks in specific sites of the genome.[112–115] This strategy has been demonstrated feasible even in primary cells.[140–142] Here, we have applied the CRISPR/Cas9 system for this purpose.

The use of genome editing to model diseases can also resemble specific mutations found in patients. This fact will facilitate the study of specific mutations and the potential genotype-phenotype association. Herein, we have focused our genome editing events on mutations found in PKD patients.[143] Exons 8, 9 and 11 were studied since these exons have the highest mutations burden in PKD patients, being the tools developed for exon 9 the

most efficient in terms of gene editing. In fact, all treated cells, selected by the expression of GFP, present in the construct expressing the different elements of the CRISPR/Cas9 system, were properly edited in the selected genome site. This efficient gene editing was obtained in different stablished cell lines as well as in primary human hematopoietic progenitors ($CD34^+$), showing this strategy optimal and closer to patient cells.

Again, a significant toxicity was observed in primary cells when these tools were used. The most common and less toxic delivery method used to introduce CRISPR/Cas9 is electroporating as ribonucleoprotein.[144] Unfortunately, this strategy did not permitt the enrichment of the edited cells having a mix of wt and PKD-like cells at the end of the experiments insufficient to be considered a good model. The use of DNA as delivery method was conditioned by the necessity to select edited cells by GFP because modified cells cannot be phenotypically distinguished from wt cells at the end of the differentiation. A possible solution for this could be the use of the recently commercialized Cas9-GFP fusion proteins. If sorting of cells electroporated with these ribonucleoprotein complexes really represents edited cells, it will solve both problems.

The use of guides separately resulted in the classical indel distribution around cut site.[116] However, the use of guides in pairs turned out as a good strategy to control the events generated. This strategy has already been used by others.[145–147] In our case, after using two guides at the same time, about 70% events generated were the expected precise deletion, something correlating with some reported data[145], but almost twice the obtained by others.[146,147] This idea of controlling events was especially interesting in the combination of guide I+II (also named combo 2) that only generated two main events, something not previously reported, the expected 31 bp deletion and a very similar deletion of 32 bp. Both events generated premature stop codons, making this combination perfect to have a PKD-like model of predictable genotype. Also, the generation of a BstXI restriction site was useful to follow the presence of the precise deletion. However, precise deletion was not possible in Combo 1 (8 bp deletion) because the sequences of both guides overlapped, thus cleavage activity of guide I altered recognition site of the other guide.

Other possible outcomes described in the literature are the activity of both guides of each couple acting separately but not generating the deletion, named here as single guide indel (*SGI*), or acting at the same time generating inversions or duplications.[148] All these

kinds of events would be present in amplicons with similar sizes to the wild type sequence. NGS study of small deletions (such as combos 1 and 2) permitted the finding of *SGI* using the same mapping used to study the deletions. However, combos 3 to 6 generated deletions too big to use this strategy. For this, to find *SGI* in the rest of the combos, two NGS mapping were performed, one of them assuming the wild type sequence as template and the other using the theoretical deletion as template. As it has been showed, *SGI* events were found in combos 3 to 6. Additional experiments would be required to find inversions or duplications as long as NGS software cannot map sequences too different from the given template.

Selection of exon 9 as gene editing target has demonstrated to be a good choice to knock out RPK, as long as all combinations tested, that start in this exon, resulted in the complete elimination of RPK protein. This means that the modification of this region is important, corroborating the *in silico* study. Also, the use of a GFP labeled Cas9 to select the cells in which gene editing tools were introduced was a highly efficient procedure.

Regarding the capacity to mature toward erythroid lineage, as commented in the previous section, the PKD patient sample was able to reach orthochromatic stage during *in vitro* differentiation, not showing cell accumulation in any immature stage that could indicate defects in maturation. This supports the idea of PKD cells differentiate *in vitro* as normal cells, without altering the erythroid surface antigen profile. PKD-like cells generated by CRISPR/Cas9 resembled perfectly this capacity and also presented the decrease of PK activity characteristic of PKD. Again, the PK activity in this differentiated cells, measured by PK/HK ratio resemble the activity found in the patient differentiated cells.

On the whole, the generation of PKD-like cells with the guide combination I+II allowed us to generate a model with a known genotype that perfectly reflects the main *in vitro* features of the PKD cells. This is a very useful tool where evaluate the feasibility of the different gene therapy and gene editing strategies to cure the disease.

Conclusions

Conclusions

1. The loss of one *PKLR* functional allele (as relatives have) decreases erythroid parameters but not enough to reach an anemic situation.
2. An increase in erythroid progenitor population (detectable by its $CD71^+$ $CD235a^+$ immunophenotype) and in reticulocytes appear in peripheral blood samples of PKD patients as a response of the bone marrow to the anemic stress.
3. A small, although not significant, increase of hematopoietic progenitors is observed in peripheral blood of PKD patients.
4. $CD34^+$ hematopoietic cells from PKD patients do not show impairments in progenitor's content or in their differentiation ability, an important feature because this cell type will be the source for the production of the medicinal product for the PKD gene therapy clinical trial.
5. Mice transplanted with cells transduced with the therapeutic vector rescue from the deficient phenotype when VCN values in peripheral blood are higher than 0.3.
6. PKD-like cells generated by the transduction of $CD34^+$ hematopoietic progenitors with specific RPK-shRNA decrease PK activity but also undergo vector-induced cell growth alterations.
7. All mutations generated in K562 cell line with CRISPR/Cas9 tools along exon 9, intron 9 and exon 10 of *PKLR* gene completely eliminate RPK protein production, showing the importance of this region for enzyme generation.
8. The use of guides in couples to direct Cas9 activity generate predictable precise deletion pattern as the most represented event in cell lines and primary cells, making the use of guide pairs a good strategy to control the outcomes of the genome editing process.
9. Precise deletion is consistently recorded in all deletions sizes generated, ranging from 30 to 490 bp.
10. Human PKD-like cells generated from $CD34^+$ hematopoietic progenitors using CRISPR/Cas9 reproduce the characteristics of PKD hematopoietic cells, PK activity decreased and no alterations along *in vitro* erythroid differentiation being, therefore, an appropriated model to test new advanced therapies for PKD.

References

References

1. Doulatov, S., Notta, F., Laurenti, E. & Dick, J. E. Hematopoiesis: A human perspective. *Cell Stem Cell* **10,** 120–136 (2012).

2. Baron, M. H. Concise review: Early embryonic erythropoiesis: Not so primitive after all. *Stem Cells* **31,** 849–856 (2013).

3. Palis, J. Primitive and definitive erythropoiesis in mammals. *Front. Physiol.* **5,** 1–9 (2014).

4. Giger, K. M. & Kalfa, T. A. Phylogenetic and Ontogenetic View of Erythroblastic Islands. *BioMed Research International* **2015,** (2015).

5. Kaushansky, K. *et al. Williams hematology*. (2012).

6. Koury, M. J. Abnormal erythropoiesis and the pathophysiology of chronic anemia. *Blood Rev.* **28,** 49–66 (2014).

7. Moura, I. C., Hermine, O., Lacombe, C. & Mayeux, P. Erythropoiesis and transferrin receptors. *Curr. Opin. Hematol.* **22,** 193–8 (2015).

8. Papanikolaou, G. & Pantopoulos, K. Systemic iron homeostasis and erythropoiesis. *IUBMB Life* 1–15 (2017). doi:10.1002/iub.1629

9. Hom, J., Dulmovits, B. M., Mohandas, N. & Blanc, L. The erythroblastic island as an emerging paradigm in the anemia of inflammation. *Immunologic Research* **63,** 75–89 (2015).

10. Klei, T. R. L., Meinderts, S. M., van den Berg, T. K. & van Bruggen, R. From the Cradle to the Grave: The Role of Macrophages in Erythropoiesis and Erythrophagocytosis. *Front. Immunol.* **8,** (2017).

11. Heideveld, E. & van den Akker, E. Digesting the role of bone marrow macrophages on hematopoiesis. *Immunobiology* (2016). doi:10.1016/j.imbio.2016.11.007

12. Chasis, J. A. & Mohandas, N. Erythroblastic islands: Niches for erythropoiesis. *Blood*

112, 470–478 (2008).

13. Laboratories, A. Atlas de Hemalotología con interpretación de histogramas y escategramas. 11–93 (2002).

14. Lunt, S. Y. & Vander Heiden, M. G. Aerobic Glycolysis: Meeting the Metabolic Requirements of Cell Proliferation. *Annu. Rev. Cell Dev. Biol.* **27,** 441–464 (2011).

15. Fothergill-Gilmore, L. A. & Michels, P. A. M. Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* **59,** 105–235 (1993).

16. Wang, C. *et al.* Human erythrocyte pyruvate kinase: Characterization of the recombinant enzyme and a mutant form (R510Q) causing nonspherocytic hemolytic anemia. *Blood* **98,** 3113–3120 (2001).

17. Mattevi, a *et al.* Crystal structure of Escherichia coli pyruvate kinase type I: molecular basis of the allosteric transition. *Structure* **3,** 729–741 (1995).

18. Valentini, G. *et al.* Structure and function of human erythrocyte pyruvate kinase: Molecular basis of nonspherocytic hemolytic anemia. *J. Biol. Chem.* **277,** 23807–23814 (2002).

19. Desai, S. *et al.* Tissue-specific isoform switch and DNA hypomethylation of the pyruvate kinase PKM gene in human cancers. *Oncotarget* **5,** 1–9 (2013).

20. Kahn, A. & Marie, J. Pyruvate kinases from human erythrocytes and liver. in *Methods in Enzymology* **1,** 131–140 (1982).

21. Max-Audit, I. *et al.* Pattern of pyruvate kinase isozymes in erythroleukemia cell lines and in normal human erythroblasts. *Blood* **64,** 930–936 (1984).

22. Noguchi, T., Yamada, K., Inoue, H., Matsuda, T. & Tanaka, T. The L- and R-type isozymes of rat pyruvate kinase are produced from a single gene by use of different promoters. *J. Biol. Chem.* **262,** 14366–14371 (1987).

23. Kanno, H., Fujii, H., Hirono, A. & Miwa, S. cDNA cloning of human R-type pyruvate kinase and identification of a single amino acid substitution (Thr384----Met) affecting enzymatic stability in a pyruvate kinase variant (PK Tokyo) associated with

hereditary hemolytic anemia. *Proc. Natl. Acad. Sci. U. S. A.* **88,** 8218–21 (1991).

24. Yamada, K. *et al.* Identification and Characterization of Hepatocyte-specific Regulatory Regiosn of the Rat Pyruvate Kinase L Gene. *J. Biol. Chem.* **265,** 19885–19891 (1990).

25. Max-Audit, I., Eleouet, J. F. & Romeo, P. H. Transcriptional regulation of the pyruvate kinase erythroid-specific promoter. *J. Biol. Chem.* **268,** 5431–5437 (1993).

26. Kanno, H., Fujii, H. & Miwa, S. Structural analysis of human pyruvate kinase L-gene and identification of the promoter activity in erythroid cells. *Biochem. Biophys. Res. Commun.* **188,** 516–523 (1992).

27. Van Wijk, R. *et al.* Disruption of a novel regulatory element in the erythroid-specific promoter of the human PKLR gene causes severe pyruvate kinase deficiency. *Blood* **101,** 1596–1602 (2003).

28. Raich, N. & Romeo, P.-H. Erythroid regulatory elements. *Stem Cells* **11,** 95–104 (1993).

29. Coutinho, R. *et al.* Complex inheritance of chronic haemolytic anaemia. *British Journal of Haematology* **144,** 615–616 (2009).

30. Manco, L. *et al.* A new PKLR gene mutation in the R-type promoter region affects the gene transcription causing pyruvate kinase deficiency. *Br. J. Haematol.* **110,** 993–997 (2000).

31. A., M. *et al.* A case of congenital red cell pyruvate kinase deficiency associated with hereditary spherocytosis. *Haematologica* **96,** 102 (2011).

32. View unique variants - Mendelian genes - Leiden Open Variation Database. Available at: https://grenada.lumc.nl/LOVD2/mendelian_genes/variants.php?select_db=PKLR&action=view_unique. (Accessed: 22nd August 2018)

33. Pissard, S. *et al.* Pyruvate kinase deficiency in France: A 3-year study reveals 27 new mutations. *Br. J. Haematol.* **133,** 683–689 (2006).

34. Kanno, H. *et al.* Hereditary hemolytic anemia caused by diverse point mutations of pyruvate kinase gene found in Japan and Hong Kong. *Blood* **84,** 3505–9 (1994).

35. Zanella, A., Fermo, E., Bianchi, P., Chiarelli, L. R. & Valentini, G. Pyruvate kinase deficiency: The genotype-phenotype association. *Blood Rev.* **21,** 217–231 (2007).

36. Arya, R., Layton, D. M. & Bellingham, A. J. Hereditary red cell enzymopathies. *J. Voice* **9,** 165–175 (1995).

37. Kanno, H. *et al.* Frame shift mutation, exon skipping, and a two-codon deletion caused by splice site mutations account for pyruvate kinase deficiency. *Blood* **89,** 4213–8 (1997).

38. Wong, P. A hypothesis of haemolysis in haemolytic anaemias associated with enzymopathies. *Med. Hypotheses* **59,** 105–109 (2002).

39. Martinov, M. V., Plotnikov, A. G., Vitvitsky, V. M. & Ataullakhanov, F. I. Deficiencies of glycolytic enzymes as a possible cause of hemolytic anemia. *Biochim. Biophys. Acta - Gen. Subj.* **1474,** 75–87 (2000).

40. Tomoda, A., Lachant, N. A., Noble, N. A. & Tanaka, K. R. Inhibition of the pentose phosphate shunt by 2,3-diphosphoglycerate in erythrocyte pyruvate kinase deficiency. *Br. J. Haematol.* **54,** 475–484 (1983).

41. Koralkova, P., Van Solinge, W. W. & Van Wijk, R. Rare hereditary red blood cell enzymopathies associated with hemolytic anemia - pathophysiology, clinical aspects, and laboratory diagnosis. *International Journal of Laboratory Hematology* **36,** 388–397 (2014).

42. Mohandas, N. & Gallagher, P. G. Red cell membrane : past , present , and future. *Blood* **112,** 3939–3948 (2009).

43. Andolfo, I., Russo, R., Gambale, A. & Iolascon, A. New insights on hereditary erythrocyte membrane defects. *Haematologica* **101,** 1284–1294 (2016).

44. Bogdanova, A., Makhro, A., Wang, J., Lipp, P. & Kaestner, L. Calcium in red blood cells-a perilous balance. *Int. J. Mol. Sci.* **14,** 9848–9872 (2013).

45. Valentine, W. N., Tanaka, K. R. & Paglia, D. E. Hemolytic anemias and erythrocyte enzymopathies. *Ann. Intern. Med.* **103,** 245–257 (1985).

46. Zerez, C. R. & Tanaka, K. R. Impaired nicotinamide adenine dinucleotide synthesis in pyruvate kinase-deficient human erythrocytes: a mechanism for decreased total NAD content and a possible secondary cause of hemolysis. *Blood* **69,** 999–1005 (1987).

47. Brewer, G. J. 2,3-DPG and erythrocyte oxygen affinity. *Annu. Rev. Med.* **25,** 29–38 (1974).

48. Grace, R. F. *et al.* Erythrocyte pyruvate kinase deficiency: 2015 status report. *Am. J. Hematol.* **90,** 825–830 (2015).

49. Montllor, L. *et al.* Déficit de piruvato cinasa eritrocitaria en España: estudio de 15 casos. *Med. Clin. (Barc).* **148,** 23–27 (2017).

50. Canu, G., De Bonis, M., Minucci, A. & Capoluongo, E. Red blood cell PK deficiency: An update of PK-LR gene mutation database. *Blood Cells, Mol. Dis.* **57,** 100–109 (2016).

51. Grace, R. F. *et al.* The clinical spectrum of pyruvate kinase deficiency: data from the Pyruvate Kinase Deficiency Natural History Study. *Blood* blood-2017-10-810796 (2018). doi:10.1182/blood-2017-10-810796

52. Koury, M. J. & Rhodes, M. How to approach chronic anemia. *Hematology Am. Soc. Hematol. Educ. Program* **2012,** 183–90 (2012).

53. Ferreira, P. *et al.* Hydrops fetalis associated with erythrocyte pyruvate kinase deficiency. *Eur. J. Pediatr.* **159,** 481–2 (2000).

54. Mentzer, W. C., Baehner, R. L., Schmidt-Schönbein, H., Robinson, S. H. & Nathan, D. G. Selective reticulocyte destruction in erythrocyte pyruvate kinase deficiency. *J. Clin. Invest.* **50,** 688–699 (1971).

55. Nathan, D. G., Oski, F. A., Miller, D. R. & Gardner, F. H. Life-Span and Organ Sequestration of the Red Cells in Pyruvate Kinase Deficiency. *N. Engl. J. Med.* **278,**

73–81 (1968).

56. Finkenstedt, A. *et al.* Regulation of iron metabolism through GDF15 and hepcidin in pyruvate kinase deficiency. *Br. J. Haematol.* **144,** 789–793 (2009).

57. Zanella, A. *et al.* Iron status in red cell pyruvate kinase deficiency: study of Italian cases. *Br J Haematol* **83,** 485–490 (1993).

58. Zanella, A., Fermo, E., Bianchi, P. & Valentini, G. Red cell pyruvate kinase deficiency: Molecular and clinical aspects. *Br. J. Haematol.* **130,** 11–25 (2005).

59. Tanphaichitr, V. *et al.* Successful bone marrow transplantation in a child with red blood cell pyruvate kinase deficiency. *Bone Marrow Transplant.* **26,** 689–690 (2000).

60. Unal, S. & Gumruk, F. Molecular Analyses of Pyruvate Kinase Deficient Turkish Patients from a Single Center. *Pediatr. Hematol. Oncol.* **32,** 354–361 (2015).

61. Manco, L. *et al.* PK-LR gene mutations in pyruvate kinase deficient Portuguese patients. *Br. J. Haematol.* **105,** 591–595 (1999).

62. Kung, C. *et al.* AG-348 enhances pyruvate kinase activity in red blood cells from patients with pyruvate kinase deficiency. *Blood* blood-2016-11-753525 (2017). doi:10.1182/blood-2016-11-753525

63. Grace, R. F. *et al.* Effects of AG-348, a Pyruvate Kinase Activator, on Anemia and Hemolysis in Patients with Pyruvate Kinase Deficiency: Data from the DRIVE PK Study. *Blood* **128,** (2016).

64. Ayi, K. *et al.* Pyruvate kinase deficiency and malaria. *N. Engl. J. Med.* **358,** 1805–10 (2008).

65. Min-oo, G. *et al.* Pyruvate kinase deficiency in mice protects against malaria. *Nat. Genet.* **35,** 357–62 (2003).

66. Christensen, R. D., Yaish, H. M., Johnson, C. B., Bianchi, P. & Zanella, A. Six children with pyruvate kinase deficiency from one small town: Molecular characterization of the PK-LR gene. *J. Pediatr.* **159,** 695–697 (2011).

67. Kanno, H., Ballas, S. K., Miwa, S., Fujii, H. & Bowman, H. S. Molecular abnormality

of erythrocyte pyruvate kinase deficiency in the Amish. *Blood* **83,** 2311–2316 (1994).

68. Kugler, W. *et al.* Eight novel mutations and consequences on mRNA and protein level in pyruvate kinase-deficient patients with nonspherocytic hemolytic anemia. *Hum. Mutat.* **15,** 261–272 (2000).

69. Klei, T. R. L. L. *et al.* Residual pyruvate kinase activity in PKLR deficient erythroid precursors of a patient suffering from severe haemolytic anaemia. *Eur. J. Haematol.* (2017). doi:10.1111/ejh.12874

70. Diez, A. *et al.* Life-threatening nonspherocytic hemolytic anemia in a patient with a null mutation in the PKLR gene and no compensatory PKM gene expression. *Blood* **106,** 1851–1856 (2005).

71. Akiyoshi, K., Sekiguchi, K., Okamoto, T., Suenobu, S. & Izumi, T. Cord blood transplantation in a young child with pyruvate kinase deficiency. *Pediatr. Int.* **58,** 634–636 (2016).

72. Kim, M. *et al.* Hemolytic anemia with null PKLR mutations identified using whole exome sequencing and cured by hematopoietic stem cell transplantation combined with splenectomy. *Bone Marrow Transplant.* 1–4 (2016). doi:10.1038/bmt.2016.218

73. van Straaten, S. *et al.* Worldwide study of hematopoietic allogeneic stem cell transplantation in pyruvate kinase deficiency. *Haematologica* (2017). doi:10.3324/haematol.2017.177857

74. Henig, I. & Zuckerman, T. Hematopoietic stem cell transplantation-50 years of evolution and future perspectives. *Rambam Maimonides Med. J.* **5,** e0028 (2014).

75. Smetsers, S. E., Smiers, F. J., Bresters, D., Sonnevelt, M. C. & Bierings, M. B. Four decades of stem cell transplantation for Fanconi anaemia in the Netherlands. *Br. J. Haematol.* **174,** 952–961 (2016).

76. Fagioli, F. *et al.* Haematopoietic stem cell transplantation for Diamond Blackfan anaemia: A report from the Italian Association of Paediatric Haematology and Oncology Registry. *Br. J. Haematol.* **165,** 673–681 (2014).

77. Gene Therapy Clinical Trials Worldwide. Available at: http://www.abedia.com/wiley/index.html. (Accessed: 4th June 2018)

78. Tani, K. *et al.* Retrovirus-mediated gene transfer of human pyruvate kinase (PK) cDNA into murine hematopoietic cells: implications for gene therapy of human PK deficiency. *Blood* **83,** 2305–2310 (1994).

79. Meza, N. W. *et al.* Rescue of Pyruvate Kinase Deficiency in Mice by Gene Therapy Using the Human Isoenzyme. *Mol. Ther.* **17,** 2000–2009 (2009).

80. Trobridge, G. D. *et al.* Stem Cell Selection In Vivo Using Foamy Vectors Cures Canine Pyruvate Kinase Deficiency. *PLoS One* **7,** 1–10 (2012).

81. Hacein-Bey-Abina, S. LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science (80-.).* **302,** 415–419 (2003).

82. Ikawa, Y., Uchiyama, T., Jagadeesh, G. J. & Candotti, F. The long terminal repeat negative control region is a critical element for insertional oncogenesis after gene transfer into hematopoietic progenitors with Moloney murine leukemia viral vectors. *Gene Ther.* **23,** 815–818 (2016).

83. Braun, C. J. *et al.* Gene Therapy for Wiskott-Aldrich Syndrome--Long-Term Efficacy and Genotoxicity. *Sci. Transl. Med.* **6,** 227ra33-227ra33 (2014).

84. Ott, M. G. *et al.* Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat. Med.* **12,** 401–409 (2006).

85. Howe, S. J. *et al.* Insertional mutagenesis in combination with acquired somatic mutations leads to leukemogenesis following gene therapy of SCID-X1. *J. Clin* **118,** 3143–50 (2008).

86. Stein, S. *et al.* Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat. Med.* **16,** 198–204 (2010).

87. Gene Therapy Clinical Trials Worldwide. Available at:

http://www.abedia.com/wiley/vectors.php. (Accessed: 31st August 2017)

88. Garate, Z. *et al.* Generation of a High Number of Healthy Erythroid Cells from Gene-Edited Pyruvate Kinase Deficiency Patient-Specific Induced Pluripotent Stem Cells. *Stem Cell Reports* **5,** 1053–1066 (2015).

89. Chapman, B. L. & Giger, U. Inherited erythrocyte pyruvate kinase eficiency in the West Highland white terrier. *J. Small Anim. Pract.* **31,** 610–616 (1990).

90. Skelly, B. J., Wallace, M., Rajpurohit, Y. R., Wang, P. & Giger, U. Identification of a 6 base pair insertion in West Highland White Terriers with erythrocyte pyruvate kinase deficiency. *Am. J. Vet. Res.* **60,** 1169–72 (1999).

91. Harvey, J. W., Kaneko, J. J. & Hudson, E. B. Erythrocyte pyruvate kinase deficiency in a beagle dog. *Vet. Clin. Pathol.* **6,** 13–17 (1977).

92. Juvet, F. *et al.* Erythrocyte pyruvate kinase deficiency in three West Highland white terriers in Ireland and the UK. *Ir. Vet. J.* **66,** 1 (2013).

93. Inal Gultekin, G. *et al.* Erythrocytic Pyruvate Kinase Mutations Causing Hemolytic Anemia, Osteosclerosis, and Secondary Hemochromatosis in Dogs. *J. Vet. Intern. Med.* **26,** 935–944 (2012).

94. Giger, U. & Noble, N. A. Determination of erythrocyte pyruvate kinase deficiency in Basenjis with chronic hemolytic anemia. *J. Am. Vet. Med. Assoc.* **198,** 1755–61 (1991).

95. Searcy, G., Tasker, J. & Miller, D. Congenital hemolytic anemia in the Basenji dog due to erythrocyte pyruvate kinase deficiency. *Can. J. …* **35,** 67–70 (1971).

96. Whitney, K. M., Goodman, S. A., Bailey, E. M. & Lothrop, C. D. The molecular basis of canine pyruvate kinase deficiency. *Exp. Hematol.* **22,** 866–74 (1994).

97. Kohn, B. & Fumi, C. Clinical course of pyruvate kinase deficiency in Abyssinian and Somali cats. *J. Feline Med. Surg.* **10,** 145–153 (2008).

98. Grahn, R. A., Grahn, J. C., Penedo, M. C. T., Helps, C. R. & Lyons, L. A. Erythrocyte Pyruvate Kinase Deficiency mutation identified in multiple breeds of domestic cats.

BMC Vet. Res. **8,** (2012).

99. Kanno, H. *et al.* Primary structure of murine red blood cell-type pyruvate kinase (PK) and molecular characterization of PK deficiency identified in the CBA strain. *Blood* **86,** 3205–10 (1995).

100. Morimoto, M. *et al.* Pyruvate kinase deficiency of mice associated with nonspherocytic hemolytic anemia and cure of the anemia by marrow transplantation without host irradiation. *Blood* **86,** 4323–30 (1995).

101. Fortin, A. *et al.* Identification of a new malaria susceptibility locus (Char4) in recombinant congenic strains of mice. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10793–8 (2001).

102. Min-Oo, G., Tam, M., Stevenson, M. M. & Gros, P. Pyruvate kinase deficiency: Correlation between enzyme activity, extent of hemolytic anemia and protection against malaria in independent mouse mutants. *Blood Cells, Mol. Dis.* **39,** 63–69 (2007).

103. Rao, D. D., Vorhies, J. S., Senzer, N. & Nemunaitis, J. siRNA vs. shRNA: Similarities and differences. *Adv. Drug Deliv. Rev.* **61,** 746–759 (2009).

104. Wiznerowicz, M. & Trono, D. Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.* **77,** 8957–61 (2003).

105. Christofk, H. R. *et al.* The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452,** 230–233 (2008).

106. Paul, C. P., Good, P. D., Winer, I. & Engelke, D. R. Effective expression of small interfering RNA in human cells. *Nat. Biotechnol.* **20,** 505–508 (2002).

107. Eren-Koçak, E., Turner, C. A., Watson, S. J. & Akil, H. Short-hairpin RNA silencing of endogenous fibroblast growth factor 2 in rat hippocampus increases anxiety behavior. *Biol. Psychiatry* **69,** 534–540 (2011).

108. Hu, S. *et al.* Transgenic shRNA pigs reduce susceptibility to foot and mouth disease virus infection. *Elife* **4,** 1–10 (2015).

109. Rubinson, D. A. *et al.* A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat. Genet.* **33,** 401–406 (2003).

110. Kunath, T. *et al.* Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nat. Biotechnol.* **21,** 559–561 (2003).

111. Hommel, J. D., Sears, R. M., Georgescu, D., Simmons, D. L. & DiLeone, R. J. Local gene knockdown in the brain using viral-mediated RNA interference. *Nat. Med.* **9,** 1539–1544 (2003).

112. Silva, G. *et al.* Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr. Gene Ther.* **11,** 11–27 (2011).

113. Grizot, S. *et al.* Generation of redesigned homing endonucleases comprising DNA-binding domains derived from two different scaffolds. *Nucleic Acids Res.* **38,** 2006–2018 (2009).

114. Miller, J. C. *et al.* An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.* **25,** 778–85 (2007).

115. Deng, D. *et al.* Structural Basis for Sequence-Specific Recognition of. *Science (80-.).* **335,** 720–723 (2012).

116. Mojica, F. J. M. & Montoliu, L. On the Origin of CRISPR-Cas Technology: From Prokaryotes to Mammals. *Trends Microbiol.* **24,** 811–820 (2016).

117. Mojica, F. J. M. & Rodriguez-Valera, F. The discovery of CRISPR in archaea and bacteria. *FEBS J.* **283,** 3162–3169 (2016).

118. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu. Rev. Genet.* **45,** 273–297 (2011).

119. Al-Attar, S., Westra, E. R., Van Der Oost, J. & Brouns, S. J. J. Clustered regularly interspaced short palindromic repeats (CRISPRs): The hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.* **392,** 277–289 (2011).

120. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

121. Nishimasu, H. *et al.* Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162,** 1113–1126 (2015).

122. Cho, S. W., Kim, S., Kim, J. M. & Kim, J. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31,** 230–232 (2013).

123. Kingsman, S. M., Mitrophanous, K. & Olsen, J. C. Potential oncogene activity of the woodchuck hepatitis post-transcriptional regulatory element (WPRE). *Gene Ther.* **12,** 3–4 (2005).

124. Garcia-Gomez, M. *et al.* Safe and efficient gene therapy for pyruvate kinase deficiency. *Mol. Ther.* **24,** 1187–1198 (2016).

125. Optimized CRISPR Design. Available at: http://crispr.mit.edu/. (Accessed: 24th August 2018)

126. Vives Corrons, J. L. *Manual De Técnicas De Laboratorio en Hematología*. (2014). doi:10.1016/B978-84-458-2147-3.00001-X

127. Besses Raebel, C., Sans i Sabrafen, J. & Vives i Corrons, J. L. *Hematologia clínica*. (2006).

128. Merryweather-Clarke, A. Global gene expression analysis of human erythroid progenitors. *Blood* **117,** 96–108 (2011).

129. Wangen, J. R., Eidenschink Brodersen, L., Stolk, T. T., Wells, D. A. & Loken, M. R. Assessment of normal erythropoiesis by flow cytometry: Important considerations for specimen preparation. *Int. J. Lab. Hematol.* **36,** 184–196 (2014).

130. Organizacion Mundial de la Salud. Concentraciones de hemoglobina para diagnosticar la anemia y evaluar su gravedad. *VMNIS* **11.1,** 7 (2011).

131. Roy, N. B. A. *et al.* A novel 33-Gene targeted resequencing panel provides accurate, clinical-grade diagnosis and improves patient management for rare inherited anaemias. *Br. J. Haematol.* **175,** 318–330 (2016).

132. Aizawa, S. *et al.* Ineffective erythropoiesis in the spleen of a patient with pyruvate kinase deficiency. *Am. J. Hematol.* **74,** 68–72 (2003).

133. Aiuti, A. *et al.* Lentiviral hematopoietic stem cell gene therapy in patients with wiskott-aldrich syndrome. *Science (80-.).* **341,** (2013).

134. Bank, A., Dorazio, R. & Leboulch, P. A phase I/II clinical trial of β-globin gene therapy for β-thalassemia. *Ann. N. Y. Acad. Sci.* **1054,** 308–316 (2005).

135. Ronen, K. *et al.* Distribution of lentiviral vector integration sites in mice following therapeutic gene transfer to treat β-thalassemia. *Mol. Ther.* **19,** 1273–1286 (2011).

136. Cavazzana-Calvo, M. *et al.* Transfusion independence and HMGA2 activation after gene therapy of human β-thalassaemia. *Nature* **467,** 318–322 (2010).

137. Boulad, F., Mansilla-Soto, J., Cabriolu, A., Rivière, I. & Sadelain, M. Gene Therapy and Genome Editing. *Hematol. Oncol. Clin. North Am.* **32,** 329–342 (2018).

138. An, D. S. *et al.* Optimization and Functional Effects of Stable Short Hairpin RNA Expression in Primary Human Lymphocytes via Lentiviral Vectors. *Mol. Ther.* **14,** 494–504 (2006).

139. Pebernard, S. & Iggo, R. D. Determinants of interferon-stimulated gene induction by RNAi vectors. *Differentiation* **72,** 103–111 (2004).

140. Liu, Z. *et al.* Efficient CRISPR/Cas9-Mediated Versatile, Predictable, and Donor-Free Gene Knockout in Human Pluripotent Stem Cells. *Stem Cell Reports* **0,** 107–113

141. Ma, H. *et al.* Correction of a pathogenic gene mutation in human embryos. *Nature* **548,** 413–419 (2017).

142. Ousterout, D. G. *et al.* Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nat. Commun.* **6,** 6244 (2015).

143. PKLR homepage - Mendelian genes - Leiden Open Variation Database. Available at: https://grenada.lumc.nl/LOVD2/mendelian_genes/home.php?select_db=PKLR. (Accessed: 24th July 2018)

144. DeWitt, M. *et al.* Efficient Correction of the Sickle Mutation in Human Hematopoietic Stem Cells Using a Cas9 Ribonucleoprotein Complex. *bioRxiv* 036236 (2016). doi:10.1101/036236

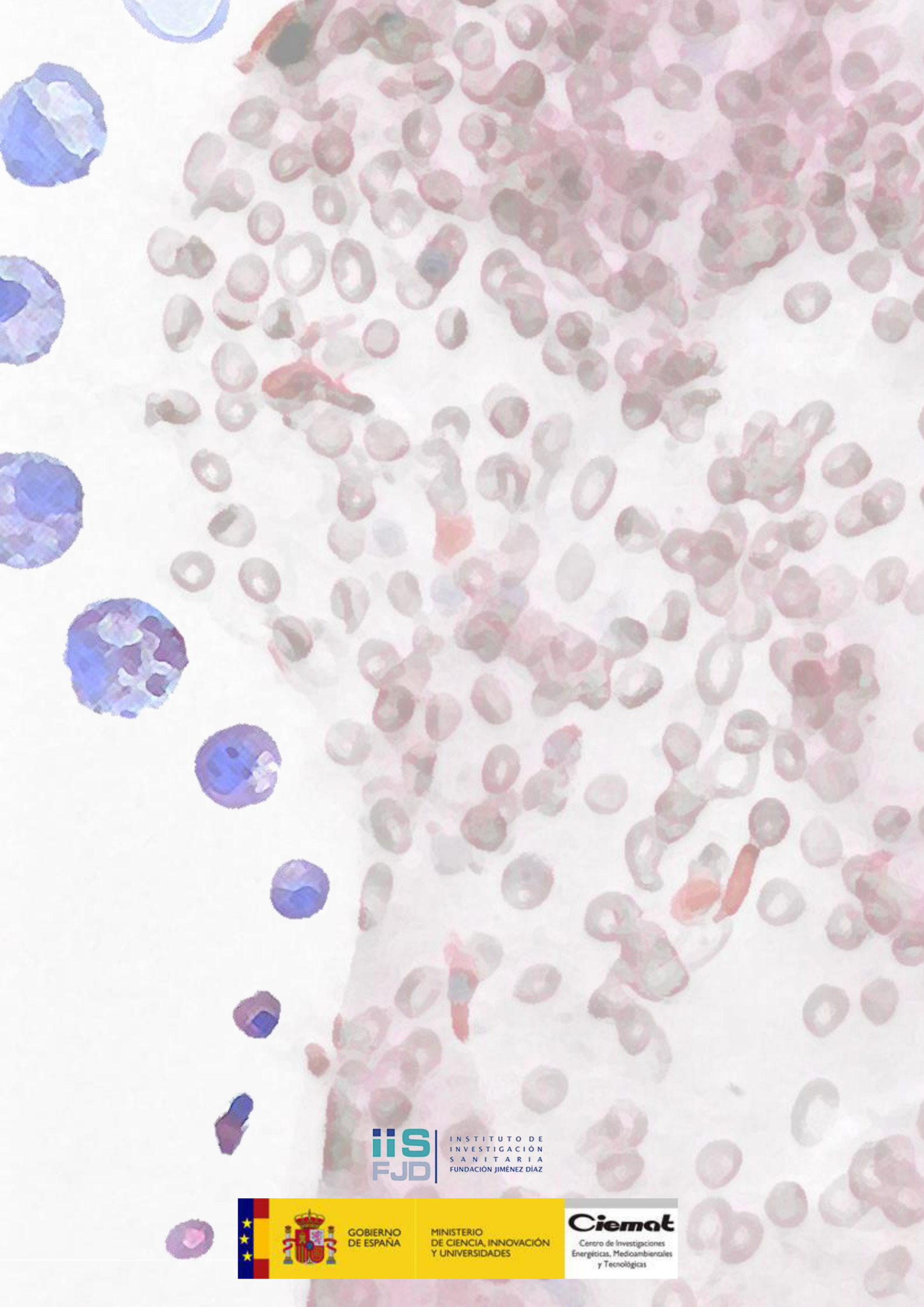
145. Geisinger, J. M., Turan, S., Hernandez, S., Spector, L. P. & Calos, M. P. In vivo blunt-end cloning through CRISPR/Cas9-facilitated non-homologous end-joining. *Nucleic Acids Res.* **44,** gkv1542 (2016).

146. Shin, H. Y. *et al.* CRISPR/Cas9 targeting events cause complex deletions and insertions at 17 sites in the mouse genome. *Nat. Commun.* **8,** 1–10 (2017).

147. Canver, M. C. *et al.* Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289,** 21312–21324 (2014).

148. Kraft, K. *et al.* Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep.* **10,** 833–839 (2015).

Appendix



INSTITUTO DE INVESTIGACIÓN SANITARIA FUNDACIÓN JIMÉNEZ DÍAZ

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Ciemat
Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas